Understanding and Mitigating Failures in Anomaly Detection: A Probabilistic Perspective

Lily H. Zhang, PhyStat 2024

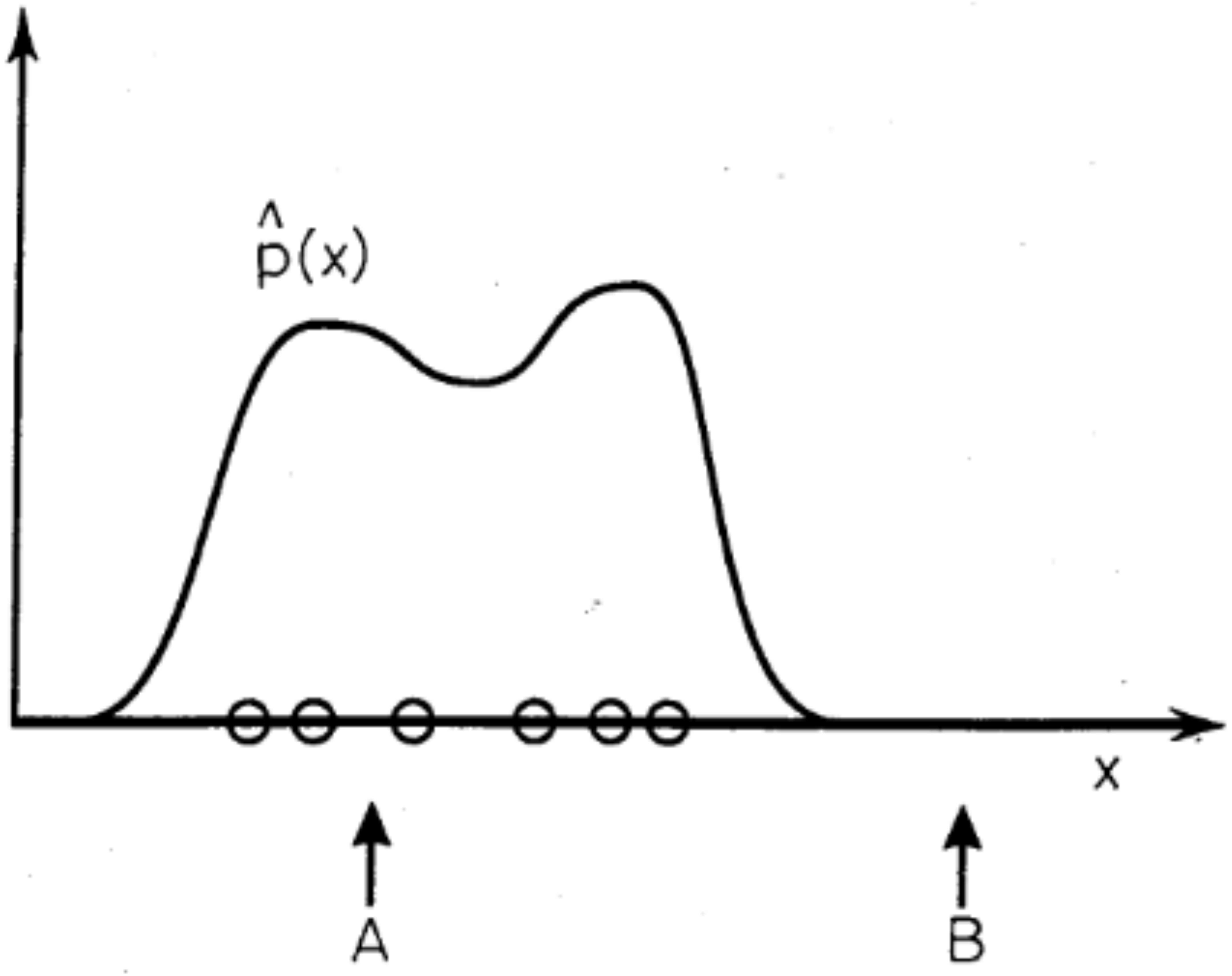
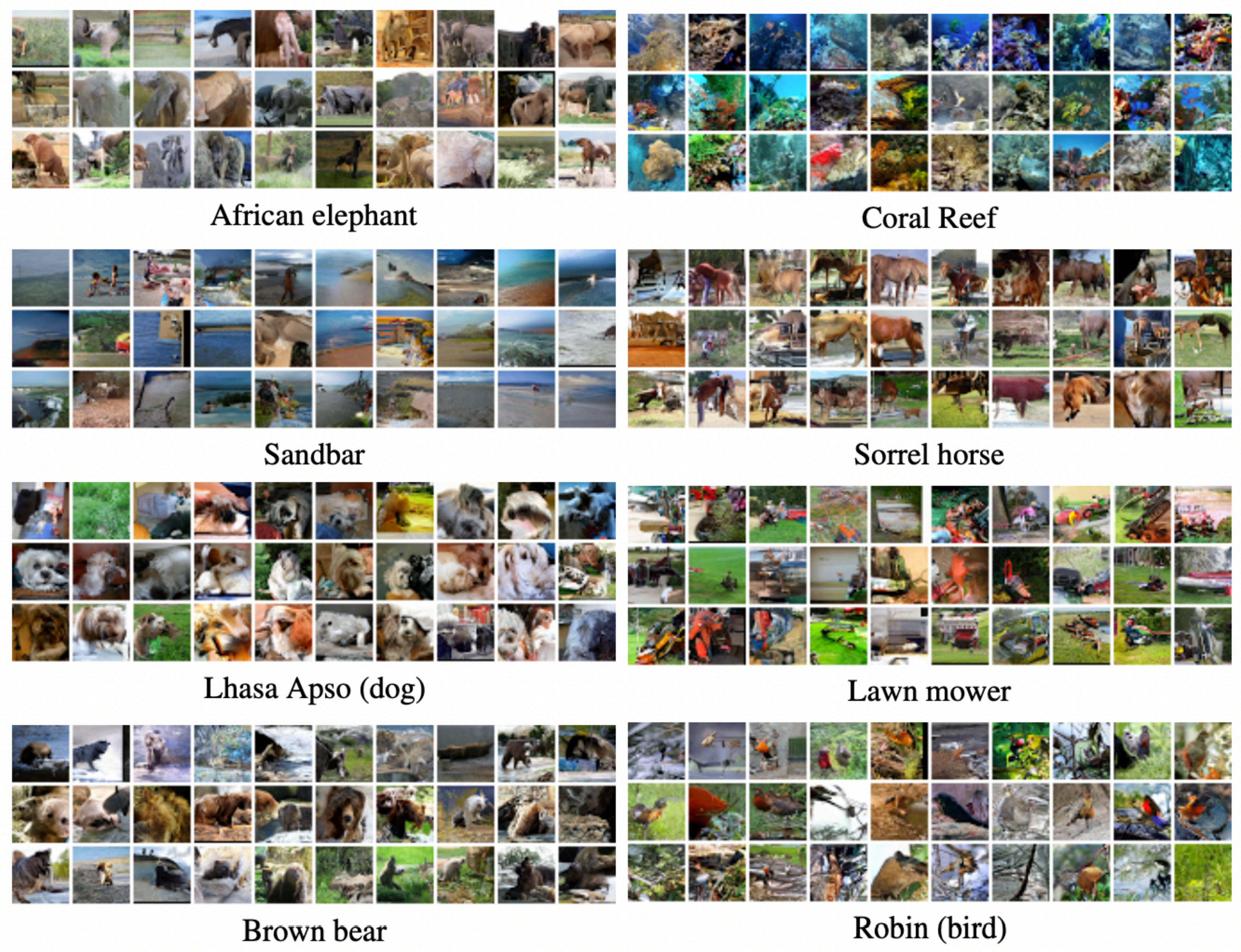


Fig. 1 *Use of the unconditional probability density to measure the degree of novelty of new input vectors*

C. M. Bishop. Novelty detection and neural network validation. *IEE Proceedings - Vision, Image and Signal Processing*, 1994.

PixelCNN



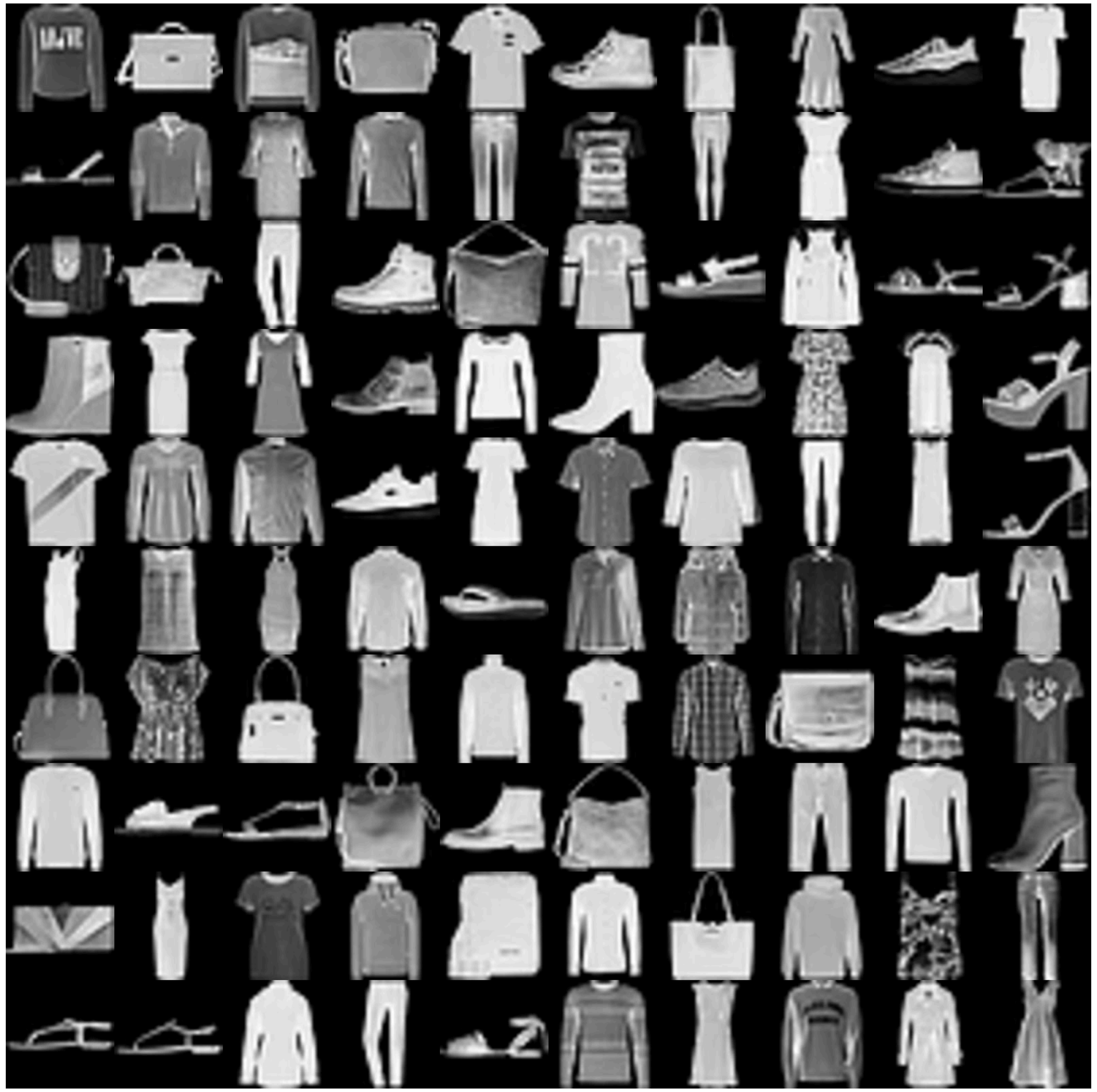
Glow

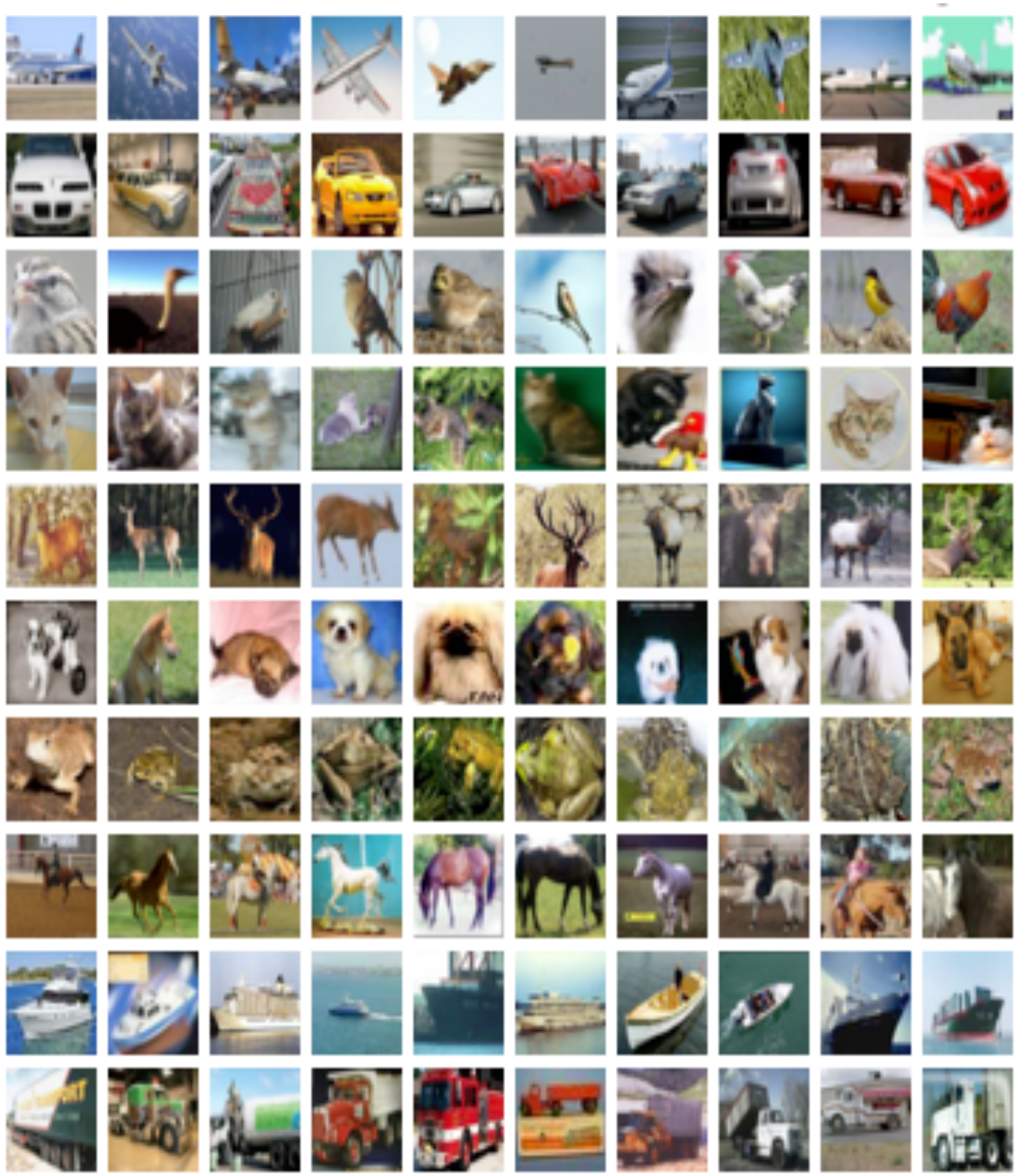
June 2016

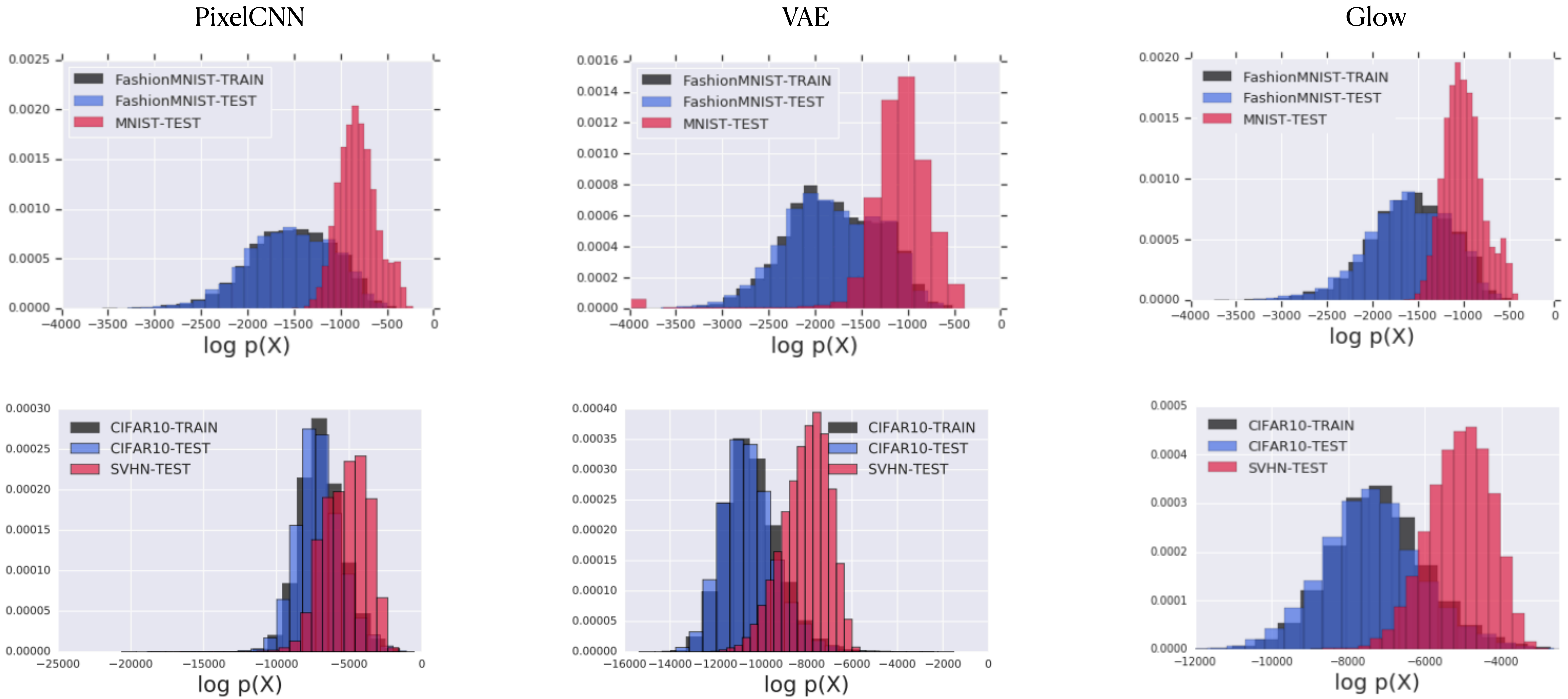
Nov 2017

July 2018

VQ-VAE







Eric Nalisnick, Akihro Matsukawa, Yee Whye Teh, Dilan Gorur, Balaji Lakshiminarayan. “Do Deep Generative Models Know What They Don’t Know?” *ICLR 2019*.

Do we need to reconsider how we perform anomaly detection?

Do we need to reconsider how we perform anomaly detection?

Understanding Anomaly Detection with Deep Invertible Networks through Hierarchies of Distributions and Features

Input Complexity and Out-of-distribution Detection with Likelihood-based Generative Models

Why Normalizing Flows Fail to Detect Out-of-Distribution Data

Article
Perfect Density Models Cannot Guarantee Anomaly Detection

Entropic Issues in Likelihood-Based OOD Detection

Do we need to reconsider how we perform anomaly detection?

Understanding Anomaly Detection with Deep Invertible Networks through Hierarchies of Distributions and Features

INPUT COMPLEXITY AND OUT-OF-DISTRIBUTION DETECTION WITH LIKELIHOOD-BASED GENERATIVE MODELS

Why Normalizing Flows Fail to Detect Out-of-Distribution Data

Article
Perfect Density Models Cannot Guarantee Anomaly Detection

Entropic Issues in Likelihood-Based OOD Detection

Likelihood Regret: An Out-of-Distribution Detection Score For Variational Auto-encoder

Likelihood Ratios for Out-of-Distribution Detection

Density of States Estimation for Out-of-Distribution Detection

DETECTING OUT-OF-DISTRIBUTION INPUTS TO DEEP GENERATIVE MODELS USING TYPICALITY

Further Analysis of Outlier Detection with Deep Generative Models

WAIC, but Why?
Generative Ensembles for Robust Anomaly Detection

Do we need to reconsider how we perform anomaly detection?

Proposition (informal): *No method can guarantee performance better than random guessing without assumptions on the out-distributions.*

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

$$\mathbf{x} \in \mathcal{X}, \phi : \mathcal{X} \to \mathbb{R}$$

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

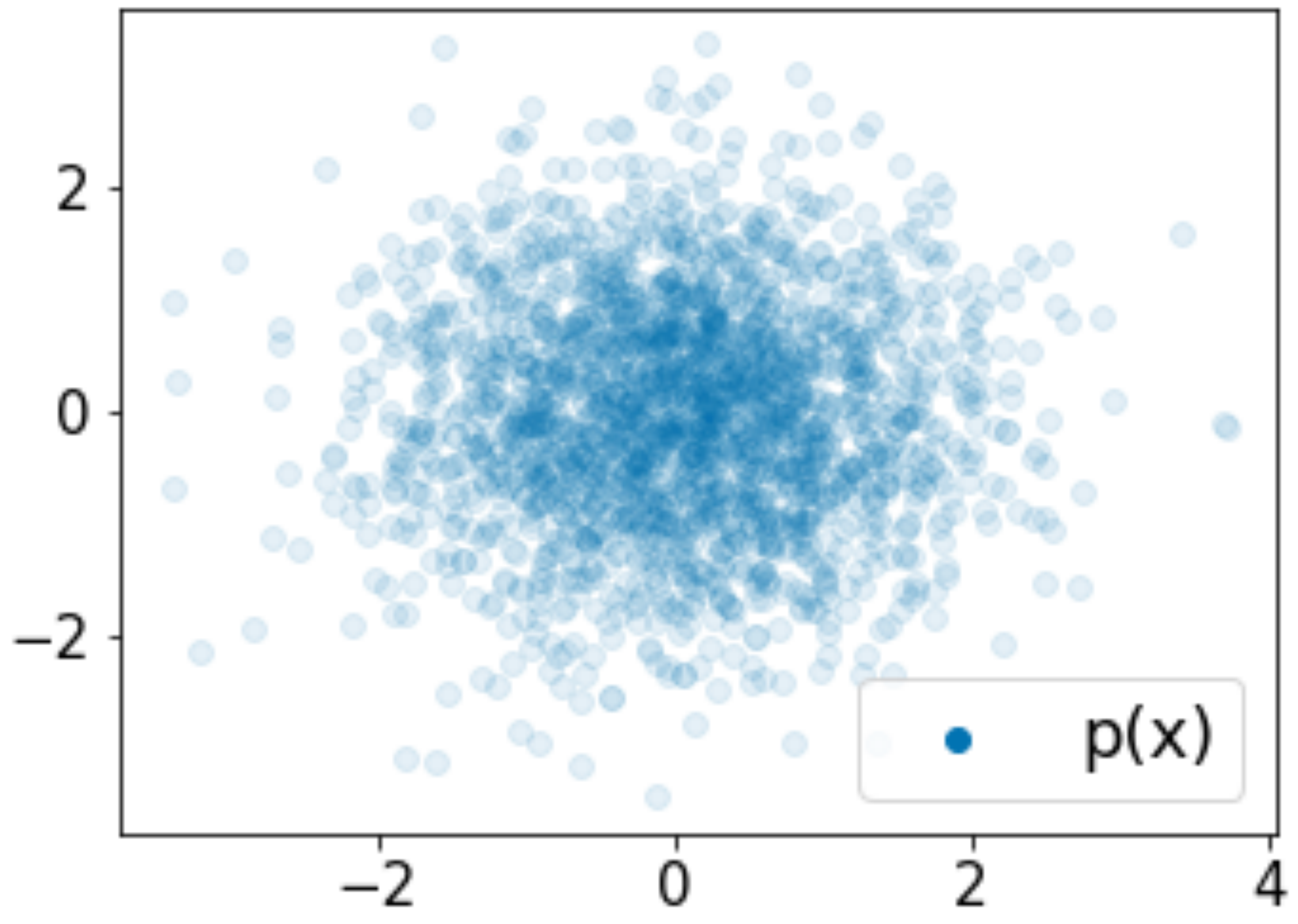
$$\mathbf{x} \in \mathcal{X}, \phi : \mathcal{X} \to \mathbb{R}$$
$$\text{e.g. } \phi_p = \log p$$

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

$$\mathbf{x} \in \mathcal{X}, \phi : \mathcal{X} \to \mathbb{R}$$
$$\text{e.g. } \phi_p = \log p$$

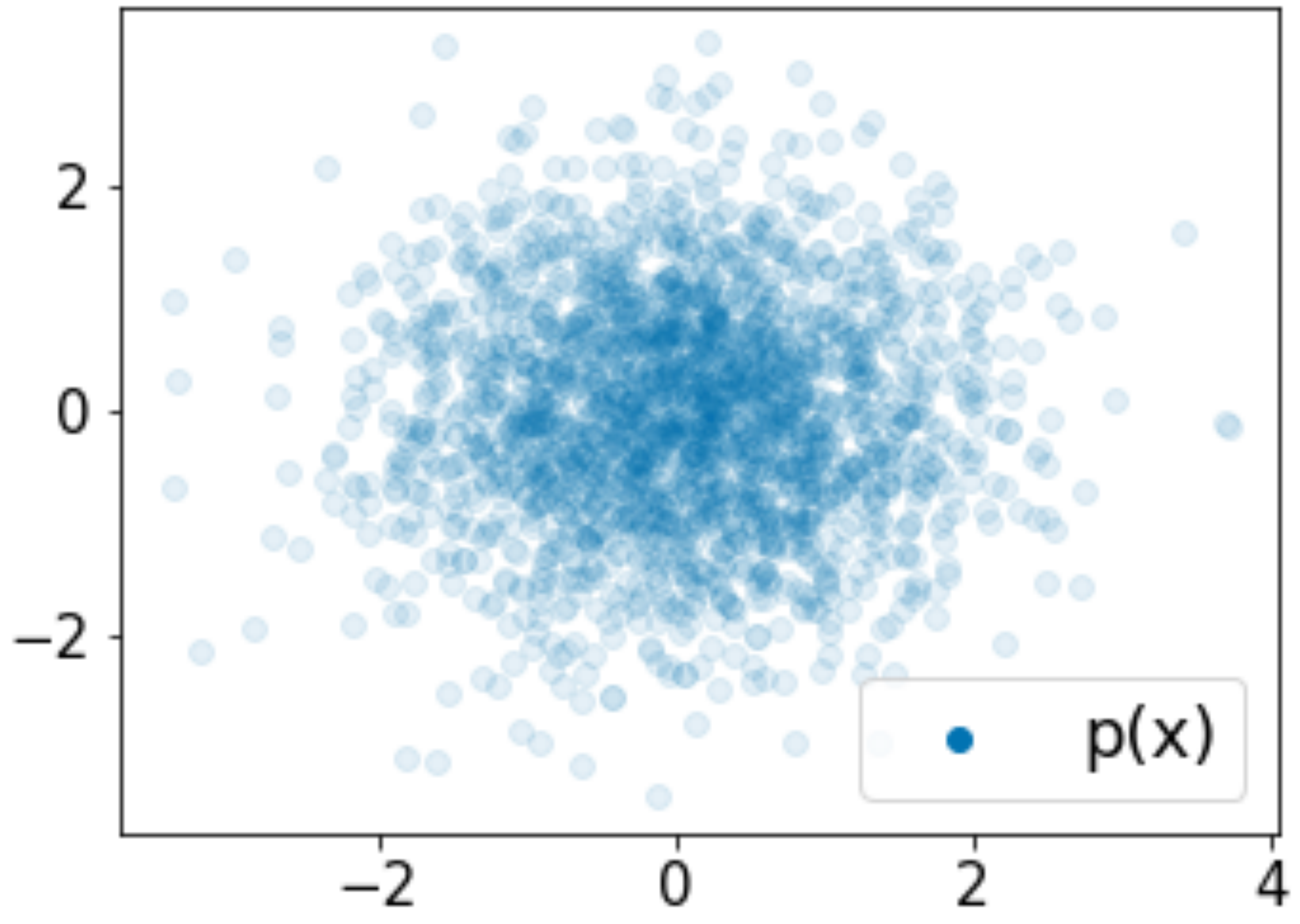


Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

$$\mathbf{x} \in \mathcal{X}, \phi : \mathcal{X} \to \mathbb{R}$$
e.g. $\phi_p = \log p$



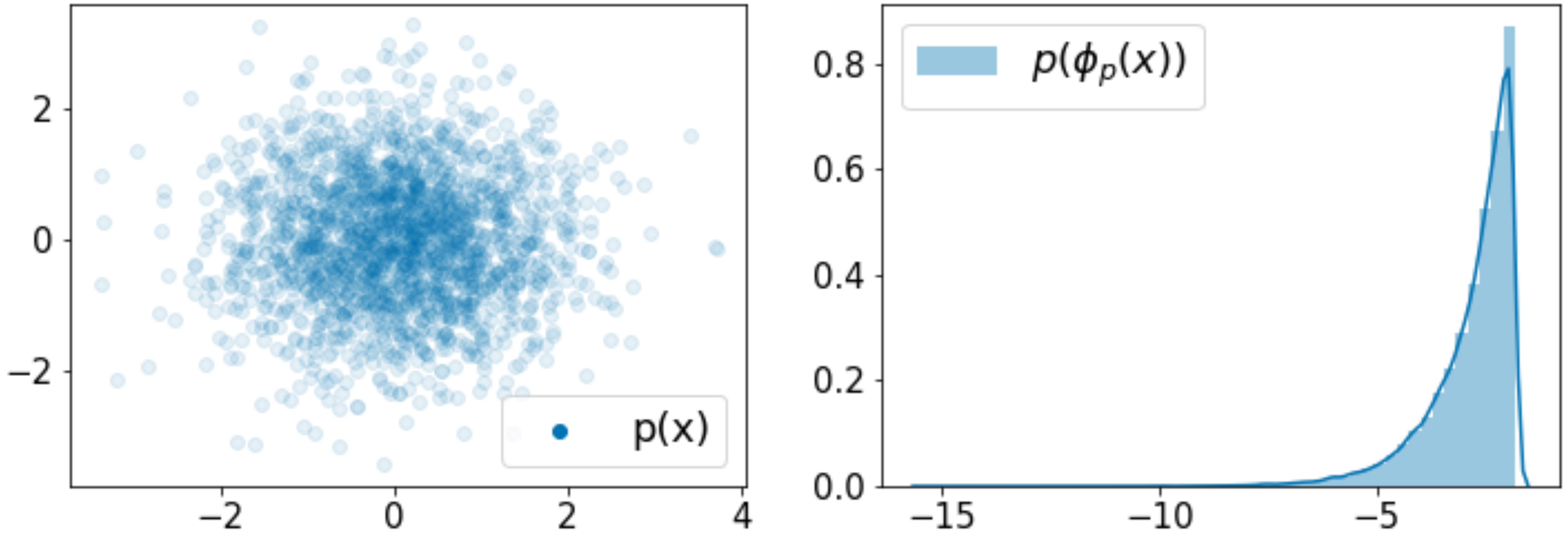
$$\phi_p = \log p$$

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

$$\mathbf{x} \in \mathcal{X}, \phi : \mathcal{X} \to \mathbb{R}$$
$$\text{e.g. } \phi_p = \log p$$

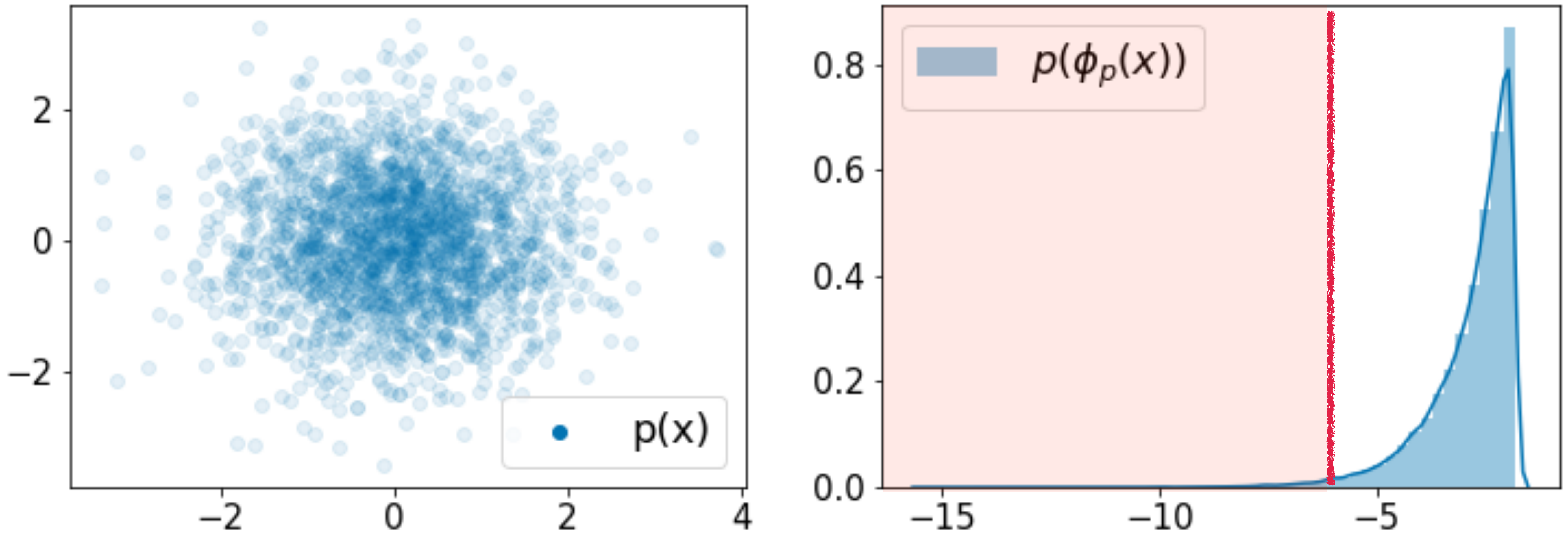


Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

$$H_0 : \mathbf{x} \sim P$$
$$H_A : \mathbf{x} \sim Q \in \mathcal{Q}, P \notin \mathcal{Q}.$$

$$\mathbf{x} \in \mathcal{X}, \phi : \mathcal{X} \to \mathbb{R}$$
$$\text{e.g. } \phi_p = \log p$$



Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

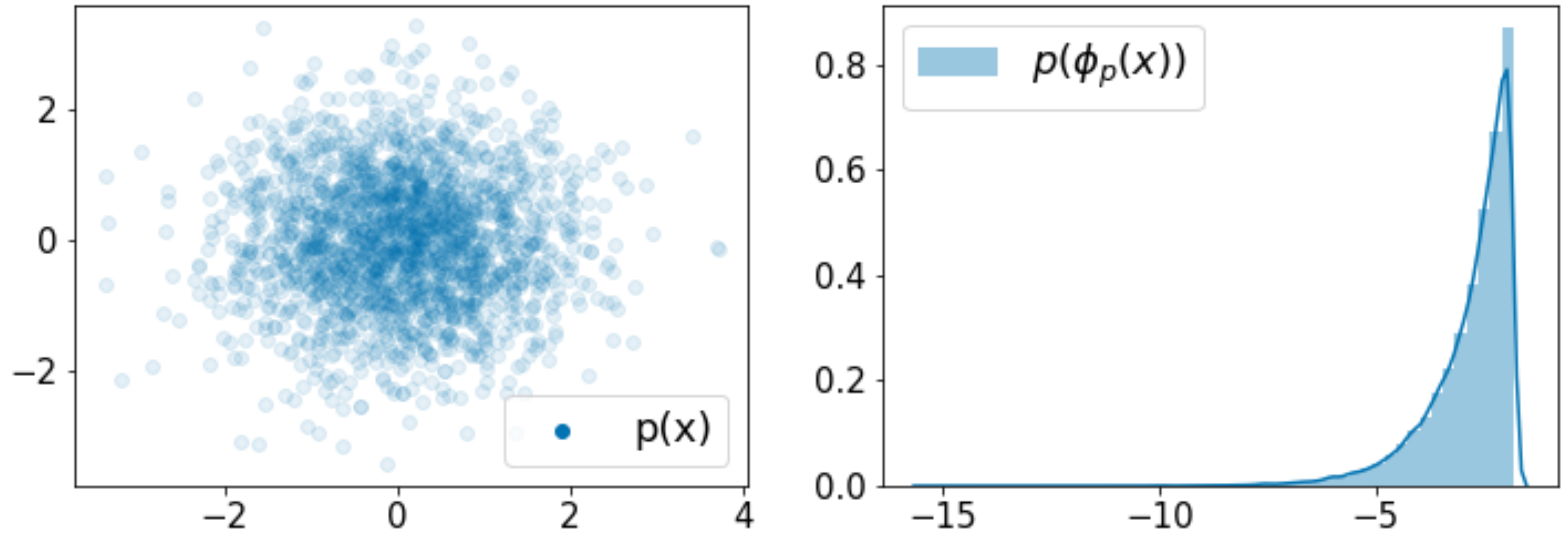
What's the right way to perform anomaly detection?

Proposition (informal): *No method can guarantee performance better than random guessing without assumptions on the out-distributions.*

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

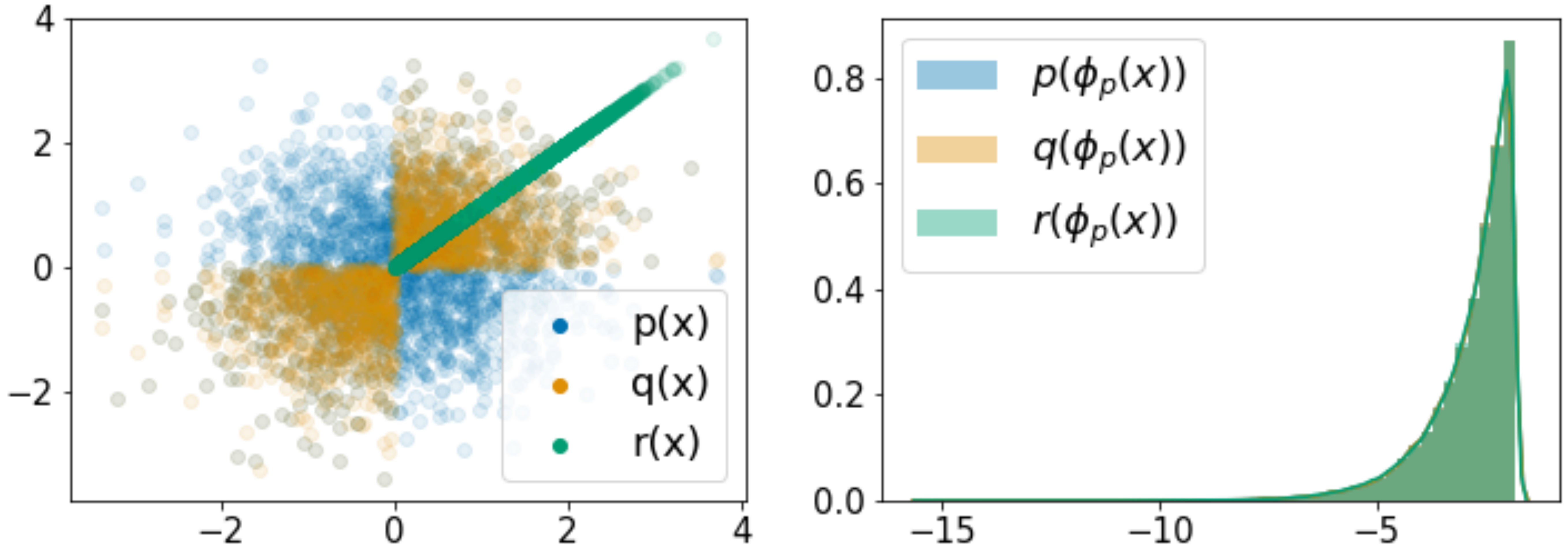
Proposition (informal): *No method can guarantee performance better than random guessing without assumptions on the out-distributions.*



Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

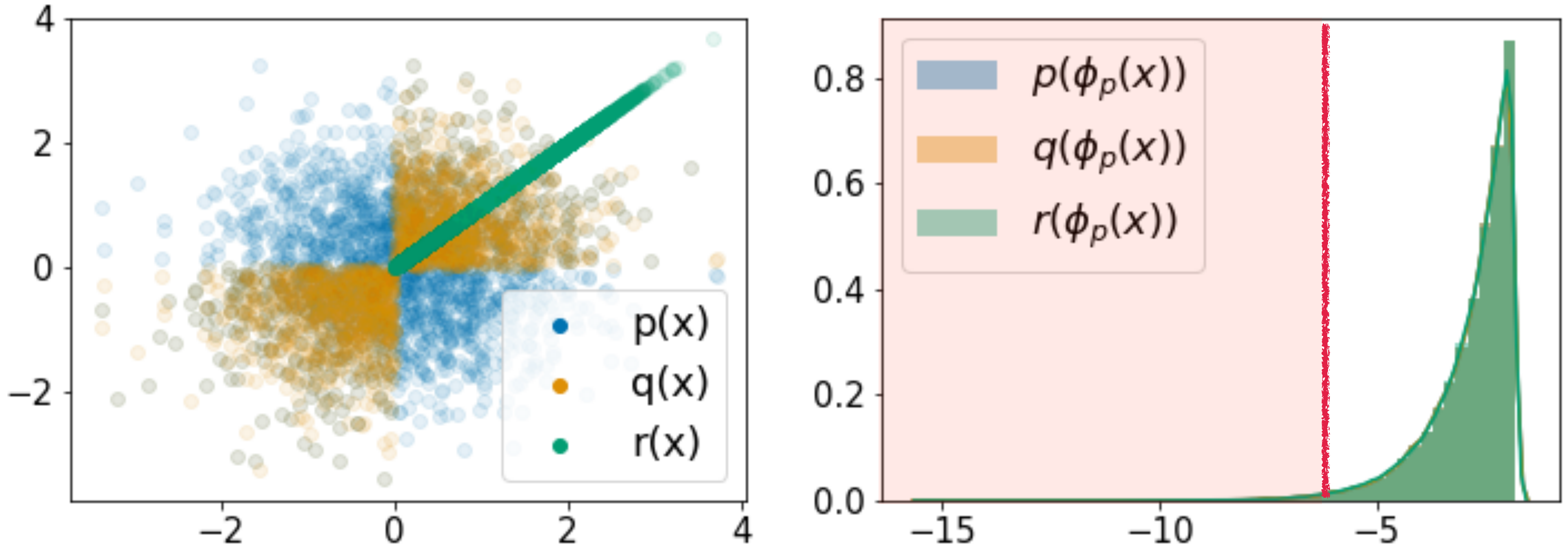
Proposition (informal): *No method can guarantee performance better than random guessing without assumptions on the out-distributions.*



Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

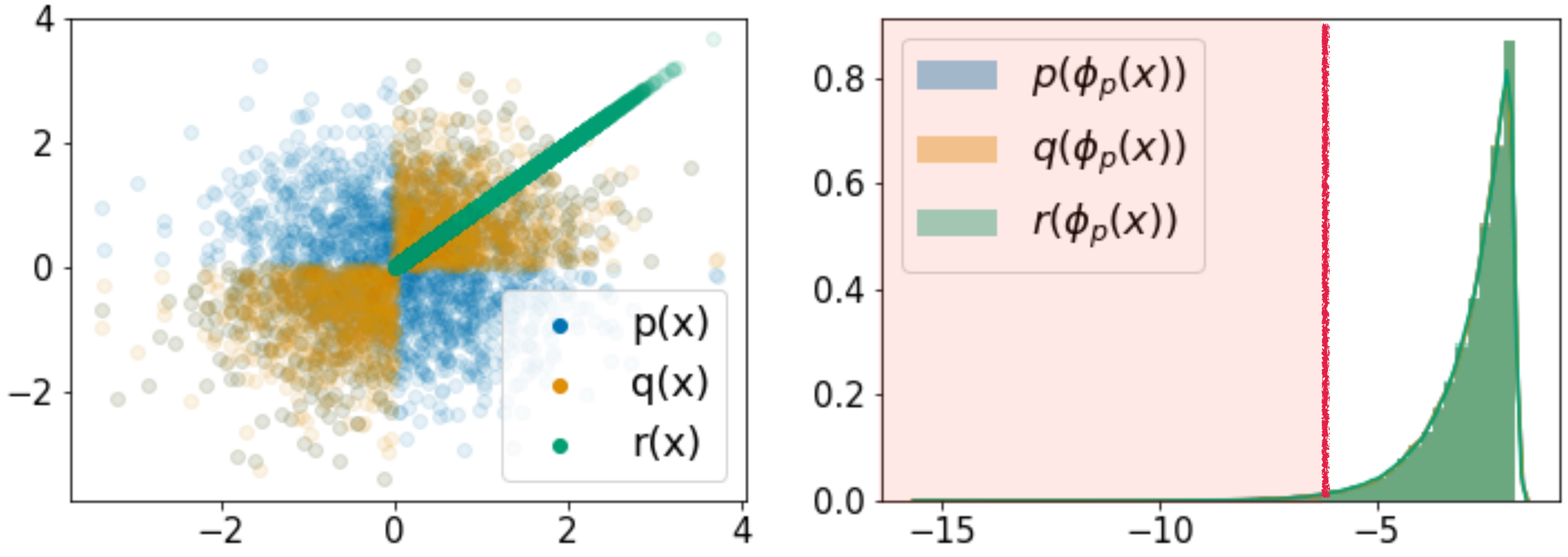
Proposition (informal): *No method can guarantee performance better than random guessing without assumptions on the out-distributions.*



Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

What's the right way to perform anomaly detection?

Proposition (informal): *No method can guarantee performance better than random guessing without assumptions on the out-distributions.*



Need to specify out-distributions of interest!

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Test statistics and their relevant out-distributions

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Test statistics and their relevant out-distributions

- Detection based on likelihood: low density under a given parametrization

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Test statistics and their relevant out-distributions

- Detection based on likelihood: low density under a given parametrization
- Detection based on likelihood ratio: low density ratio relative to a specified base distribution

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Test statistics and their relevant out-distributions

- Detection based on likelihood: low density under a given parametrization
- Detection based on likelihood ratio: low density ratio relative to a specified base distribution
- Detection based on some alternative statistic? Need to justify the definition of anomalous!

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Test statistics and their relevant out-distributions

- Detection based on likelihood: low density under a given parametrization
- Detection based on likelihood ratio: low density ratio relative to a specified base distribution
- Detection based on some alternative statistic? Need to justify the definition of anomalous!

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Test statistics and their relevant out-distributions

- Detection based on likelihood: low density under a given parametrization
- Detection based on likelihood ratio: low density ratio relative to a specified base distribution
- Detection based on some alternative statistic? Need to justify the definition of anomalous!

But sometimes alternative statistics just work well empirically…how do we reason about this?

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

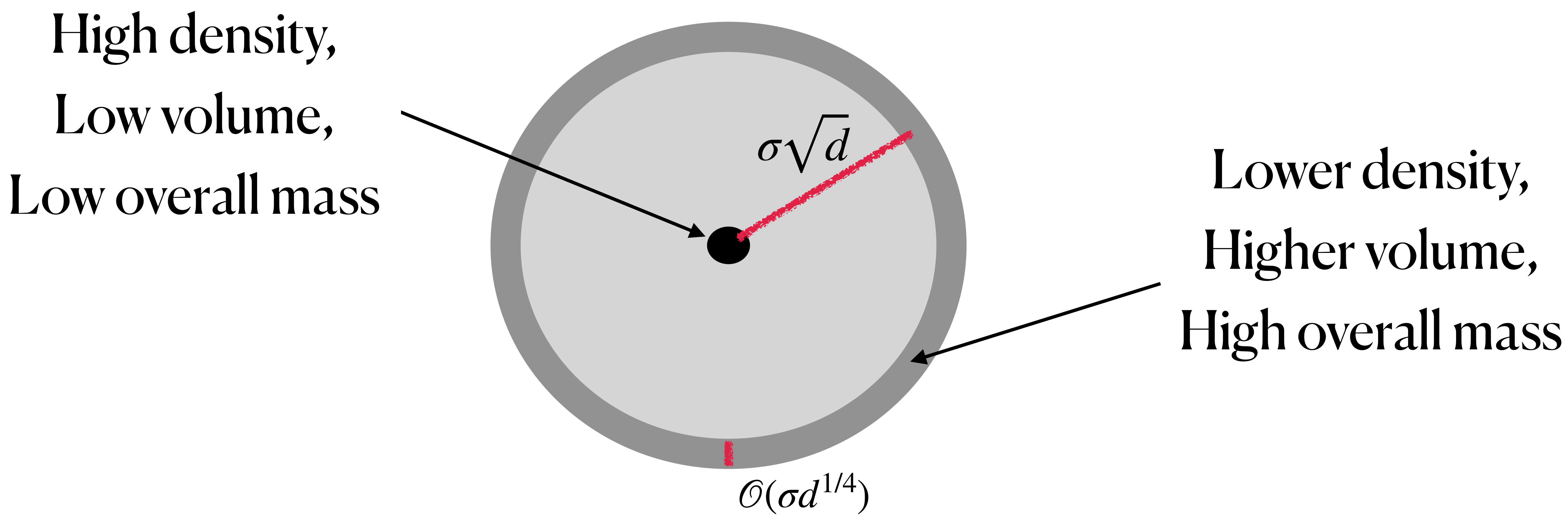
Typical set hypothesis

Typical set hypothesis

$$\phi_{\text{typical}}(\mathbf{x}) = -\left|\log p_\theta(\mathbf{x}) - \frac{1}{n}\sum_{\mathbf{x}' \in D_{tr}} \log p_\theta(\mathbf{x}')\right|$$

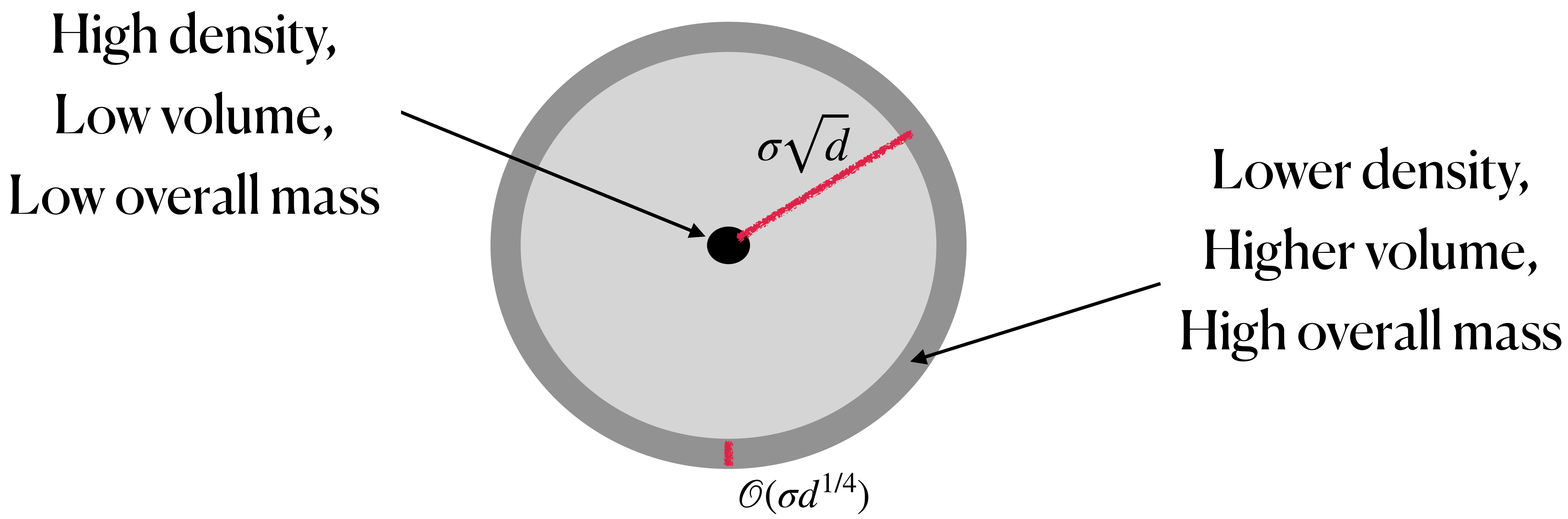
Typical set hypothesis

$$\phi_{\text{typical}}(\mathbf{x}) = -\left| \log p_\theta(\mathbf{x}) - \frac{1}{n} \sum_{\mathbf{x}' \in D_{tr}} \log p_\theta(\mathbf{x}') \right|$$



Typical set hypothesis

$$\phi_{\text{typical}}(\mathbf{x}) = -\left|\log p_\theta(\mathbf{x}) - \frac{1}{n}\sum_{\mathbf{x}' \in D_{tr}} \log p_\theta(\mathbf{x}')\right|$$

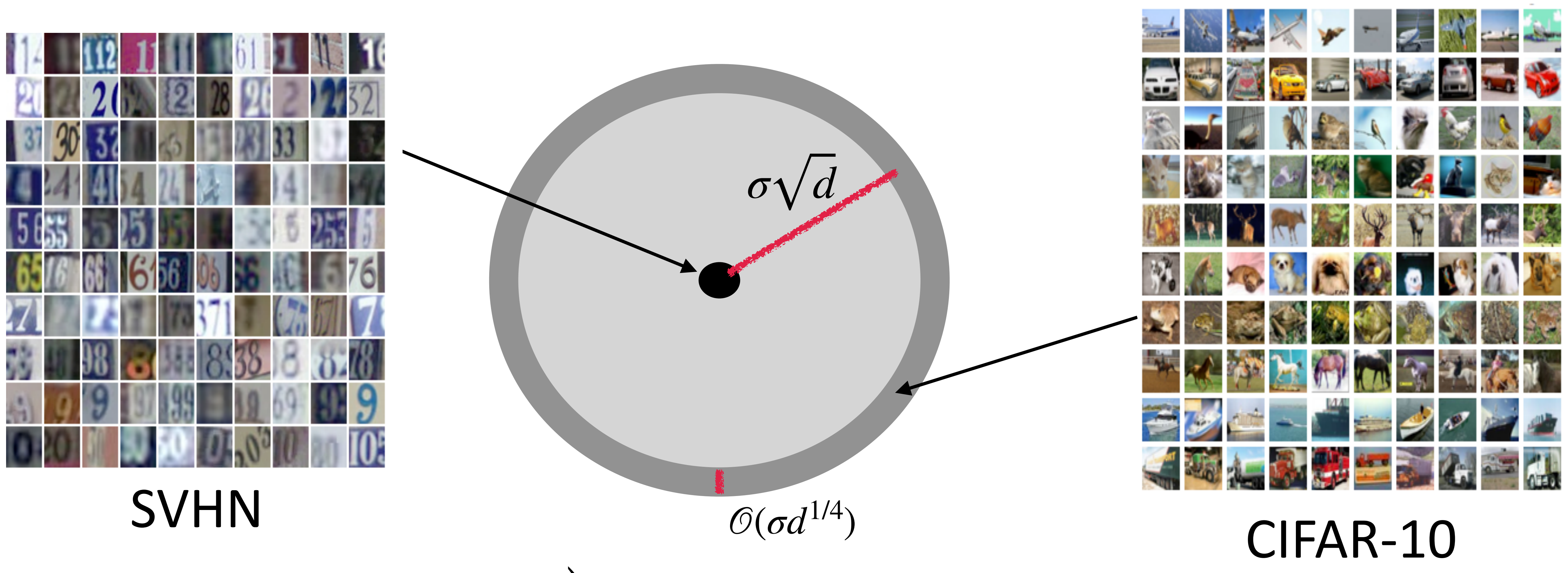


For large d:

$$P\left(\sigma\sqrt{d} - \mathcal{O}(\sigma d^{1/4}) \le |\mathbf{x}| \le \sigma\sqrt{d} + \mathcal{O}(\sigma d^{1/4})\right) \approx 1$$

Typical set hypothesis

$$\phi_{\text{typical}}(\mathbf{x}) = -\left|\log p_\theta(\mathbf{x}) - \frac{1}{n}\sum_{\mathbf{x}'\in D_{tr}} \log p_\theta(\mathbf{x}')\right|$$

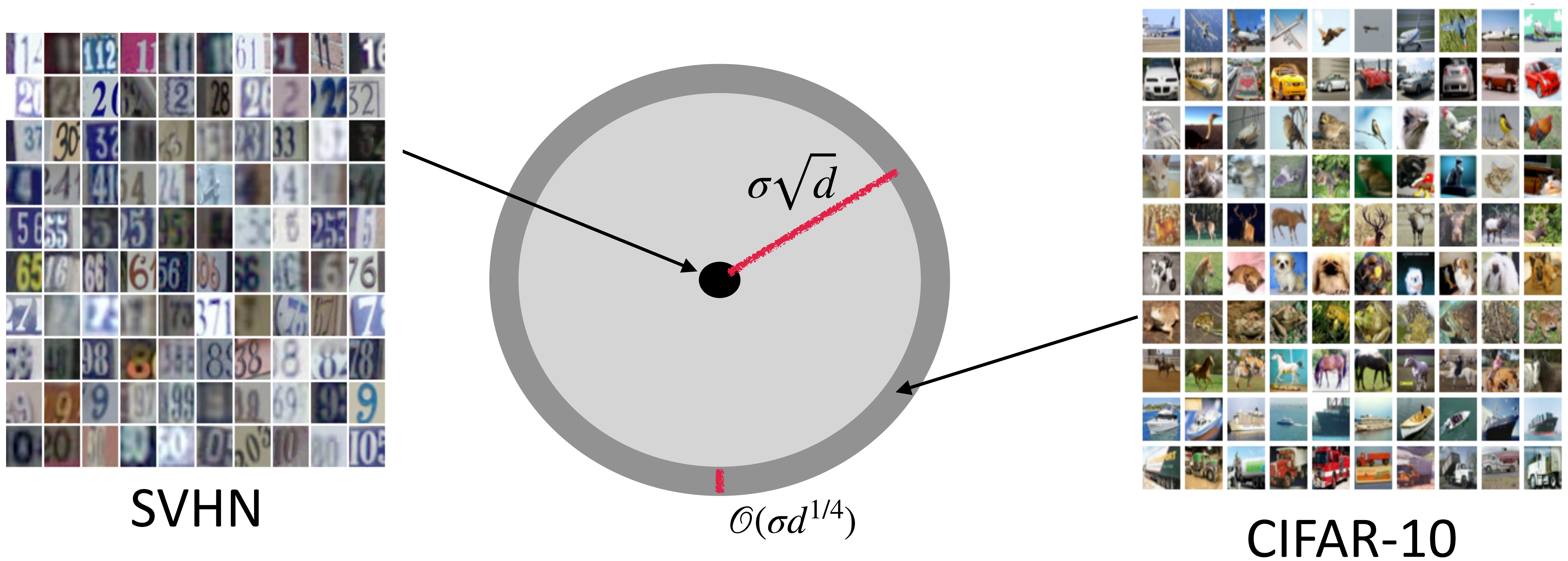


For large d:

$$P\left(\sigma\sqrt{d} - \mathcal{O}(\sigma d^{1/4}) \leq |\mathbf{x}| \leq \sigma\sqrt{d} + \mathcal{O}(\sigma d^{1/4})\right) \approx 1$$

Typical set hypothesis

$$\phi_{\text{typical}}(\mathbf{x}) = -\left|\log p_\theta(\mathbf{x}) - \frac{1}{n}\sum_{\mathbf{x}'\in D_{tr}} \log p_\theta(\mathbf{x}')\right|$$



For large d:

$$P\left(\sigma\sqrt{d} - \mathcal{O}(\sigma d^{1/4}) \leq |\mathbf{x}| \leq \sigma\sqrt{d} + \mathcal{O}(\sigma d^{1/4})\right) \approx 1$$

But the typical set assumes that relevant out-distributions overlap in support with the data distribution...

An Alternative Explanation

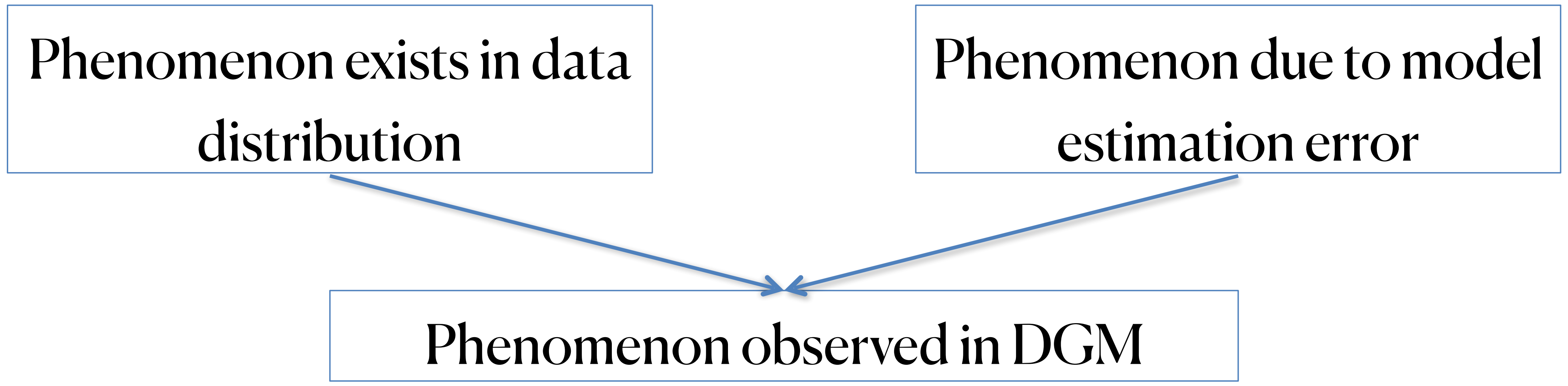
Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

An Alternative Explanation

Phenomenon observed in DGM

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

An Alternative Explanation



Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

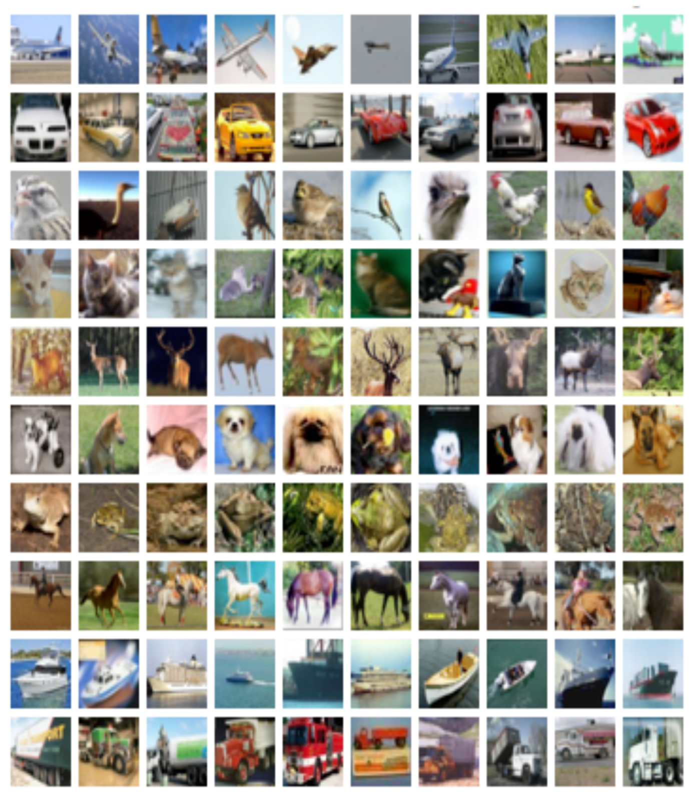
An Alternative Explanation

Phenomenon exists in data distribution

Phenomenon due to model estimation error

Phenomenon observed in DGM

SVHN



CIFAR-10

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

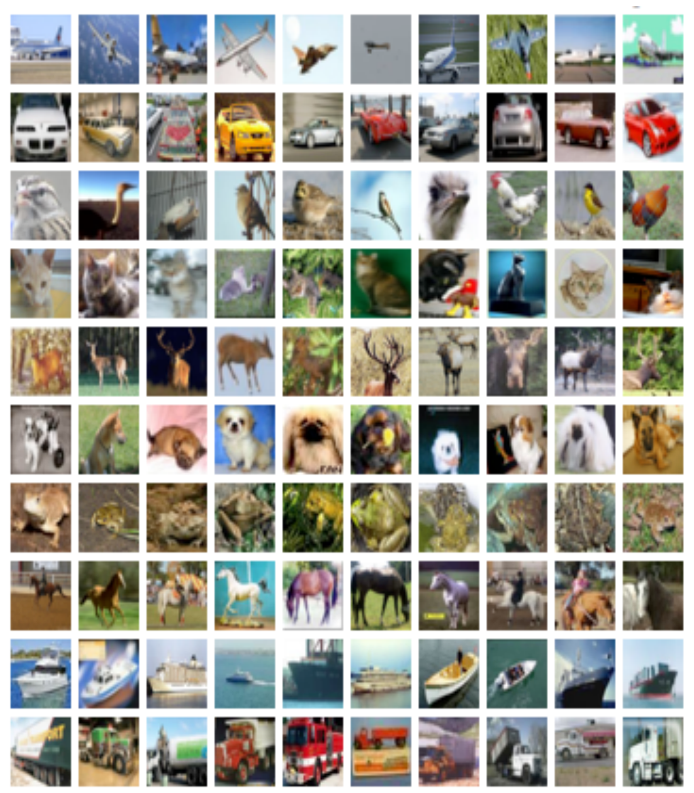
An Alternative Explanation

Phenomenon exists in data distribution

Phenomenon due to model estimation error

Phenomenon observed in DGM

SVHN



CIFAR-10

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

An Alternative Explanation

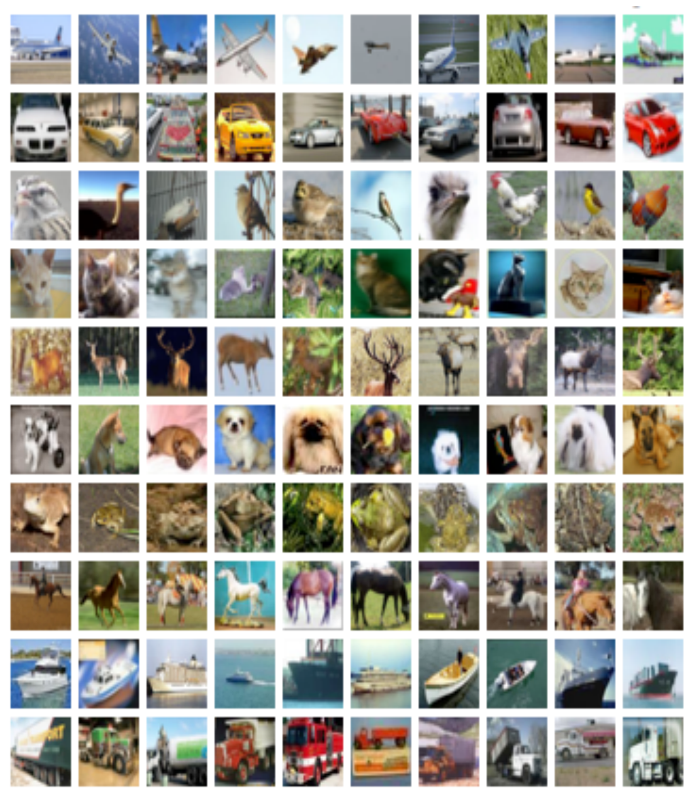
Phenomenon exists in data distribution

Phenomenon due to model estimation error

Phenomenon observed in DGM

Alternative test statistics correct for model estimation error!

SVHN



CIFAR-10

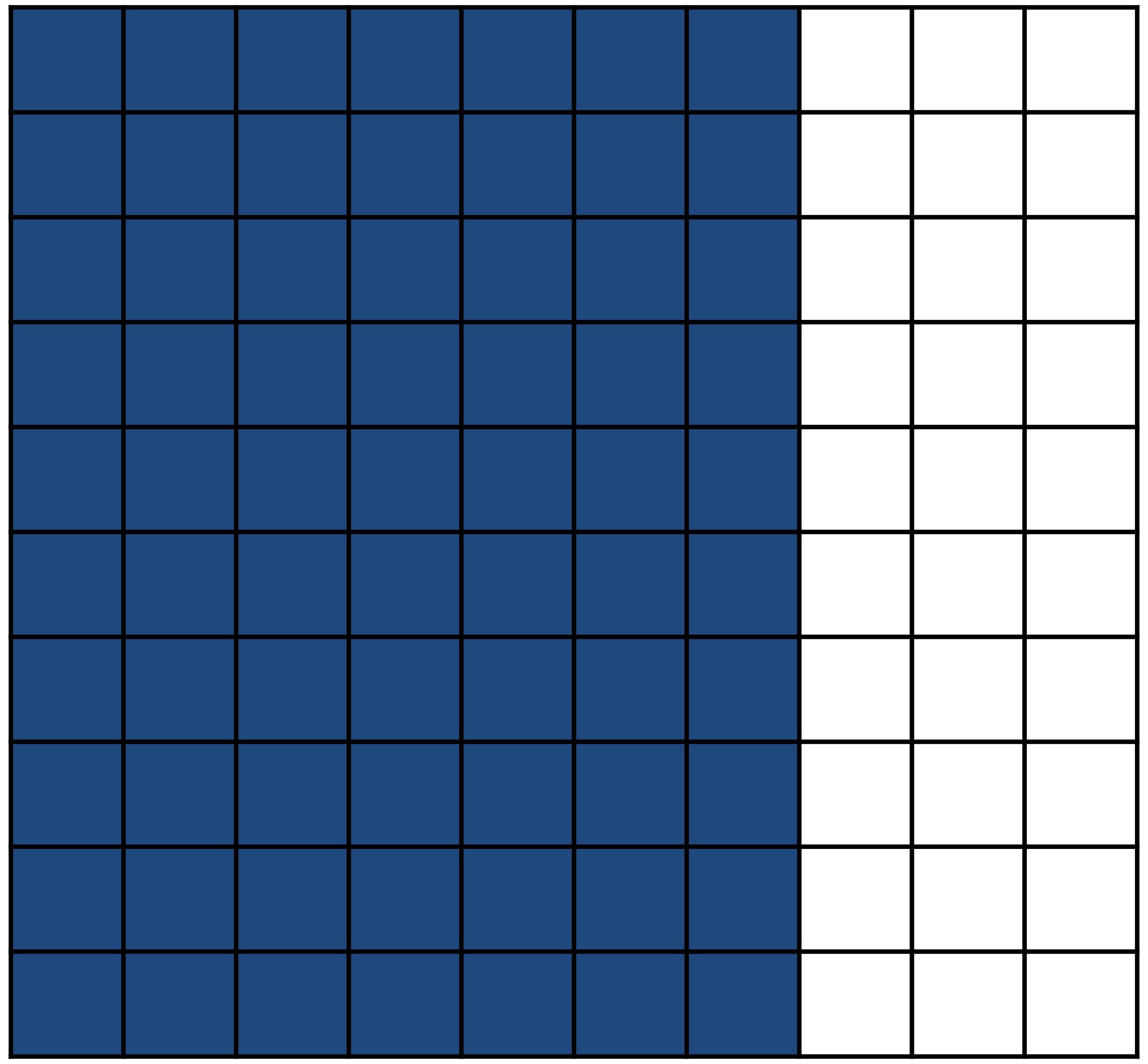
Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

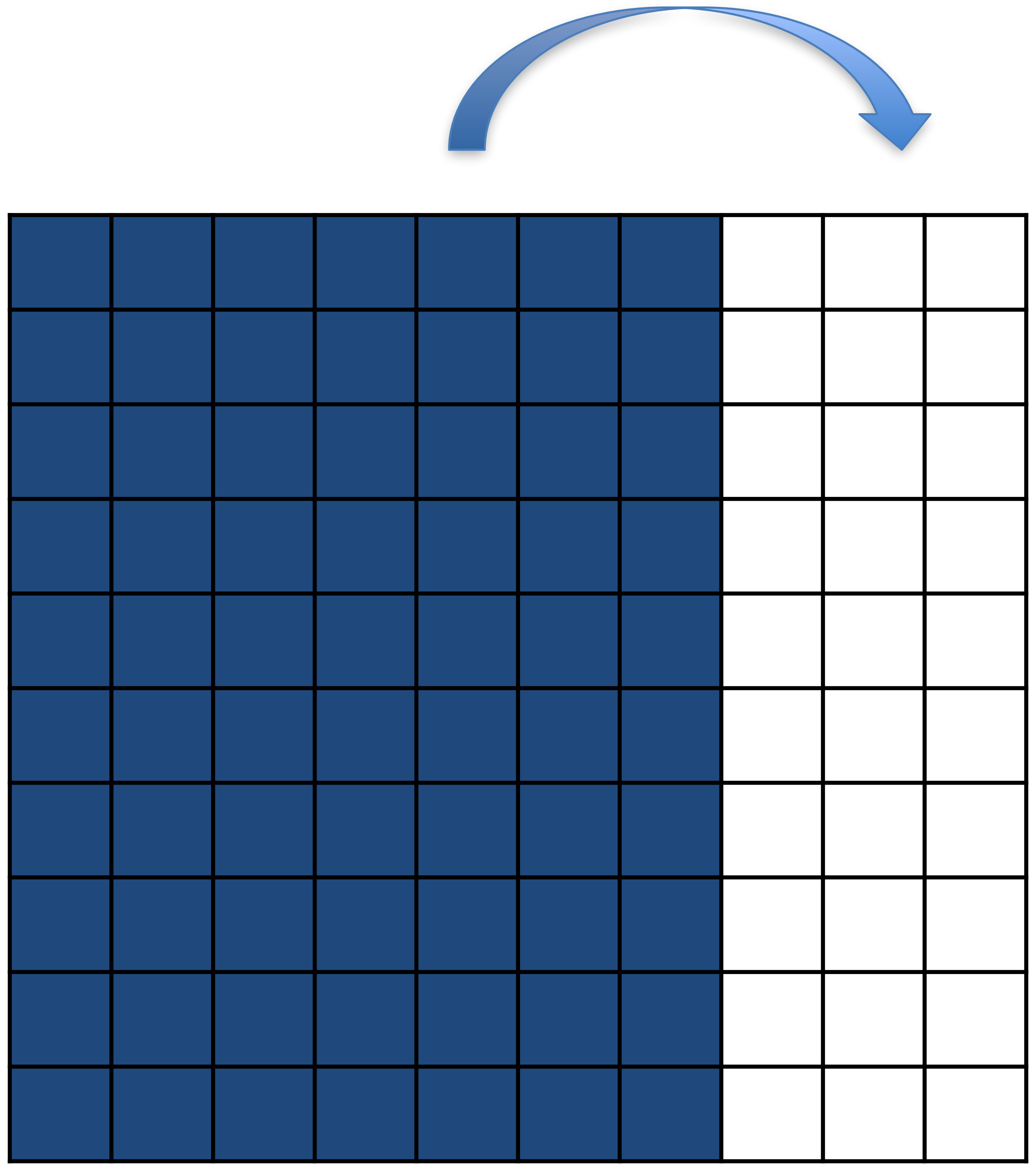
Minimal Estimation Error Can Cause Detection Failures



True distribution

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

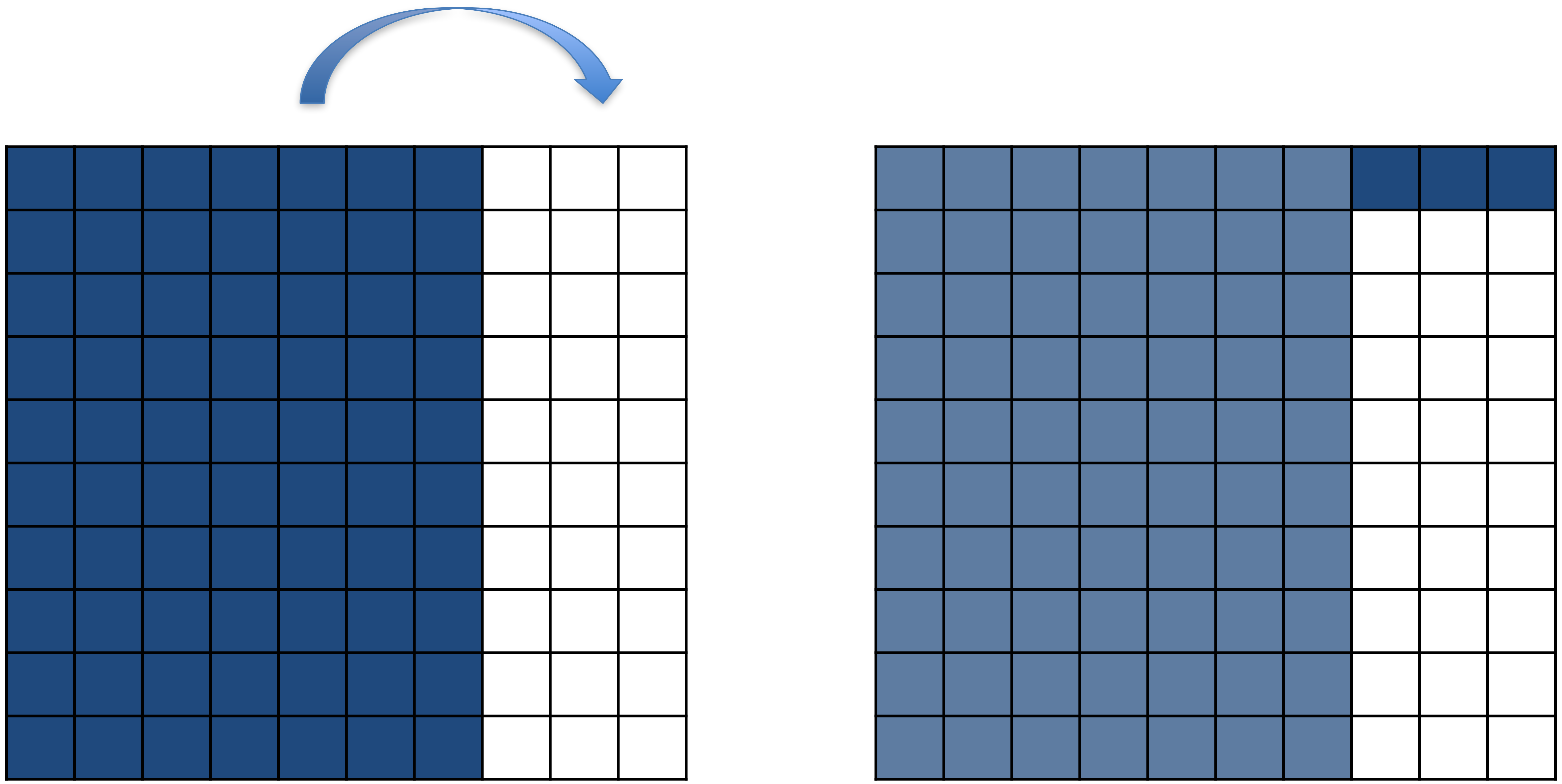
Minimal Estimation Error Can Cause Detection Failures



True distribution

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021*.

Minimal Estimation Error Can Cause Detection Failures

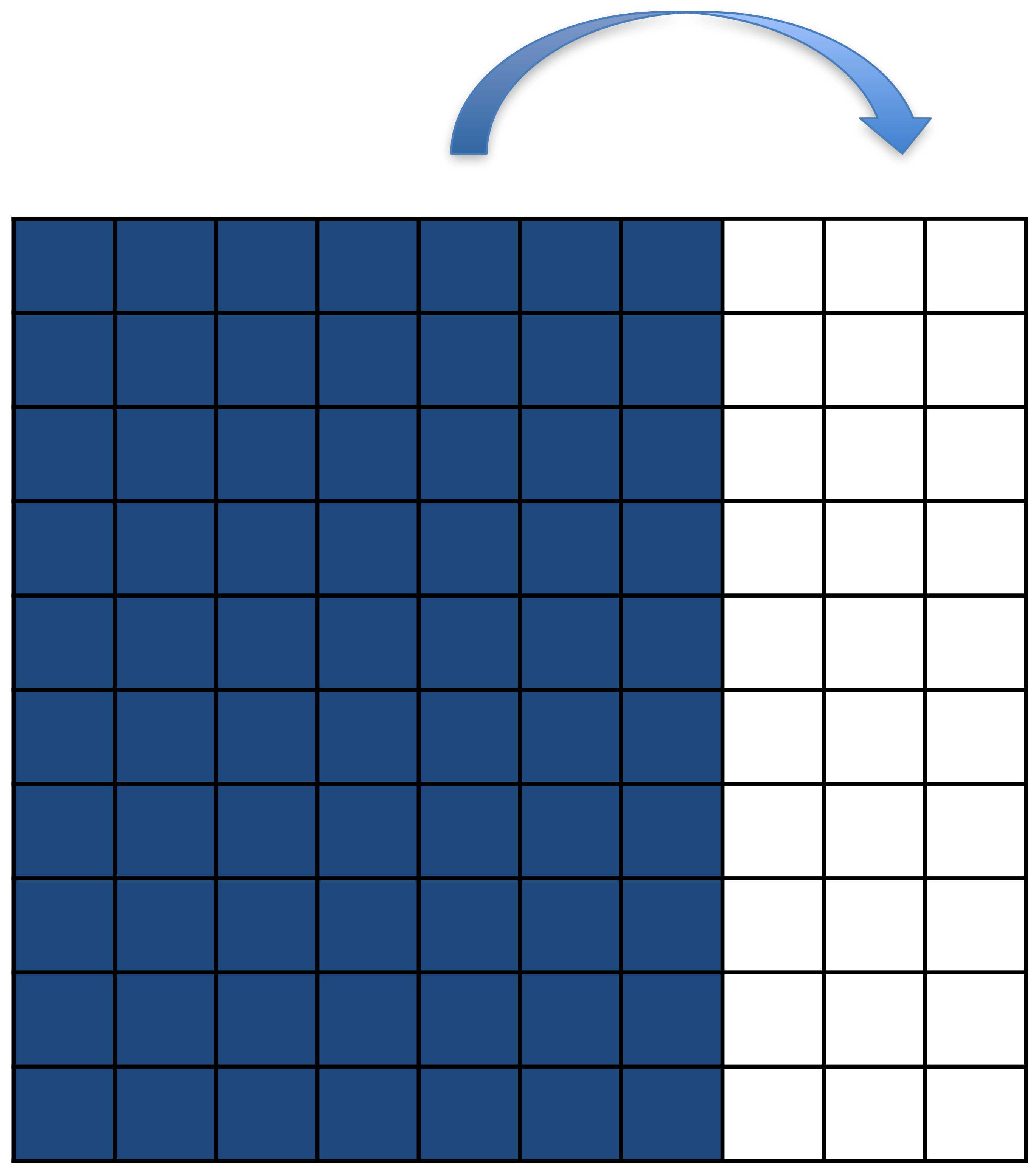


True distribution

Misestimated model

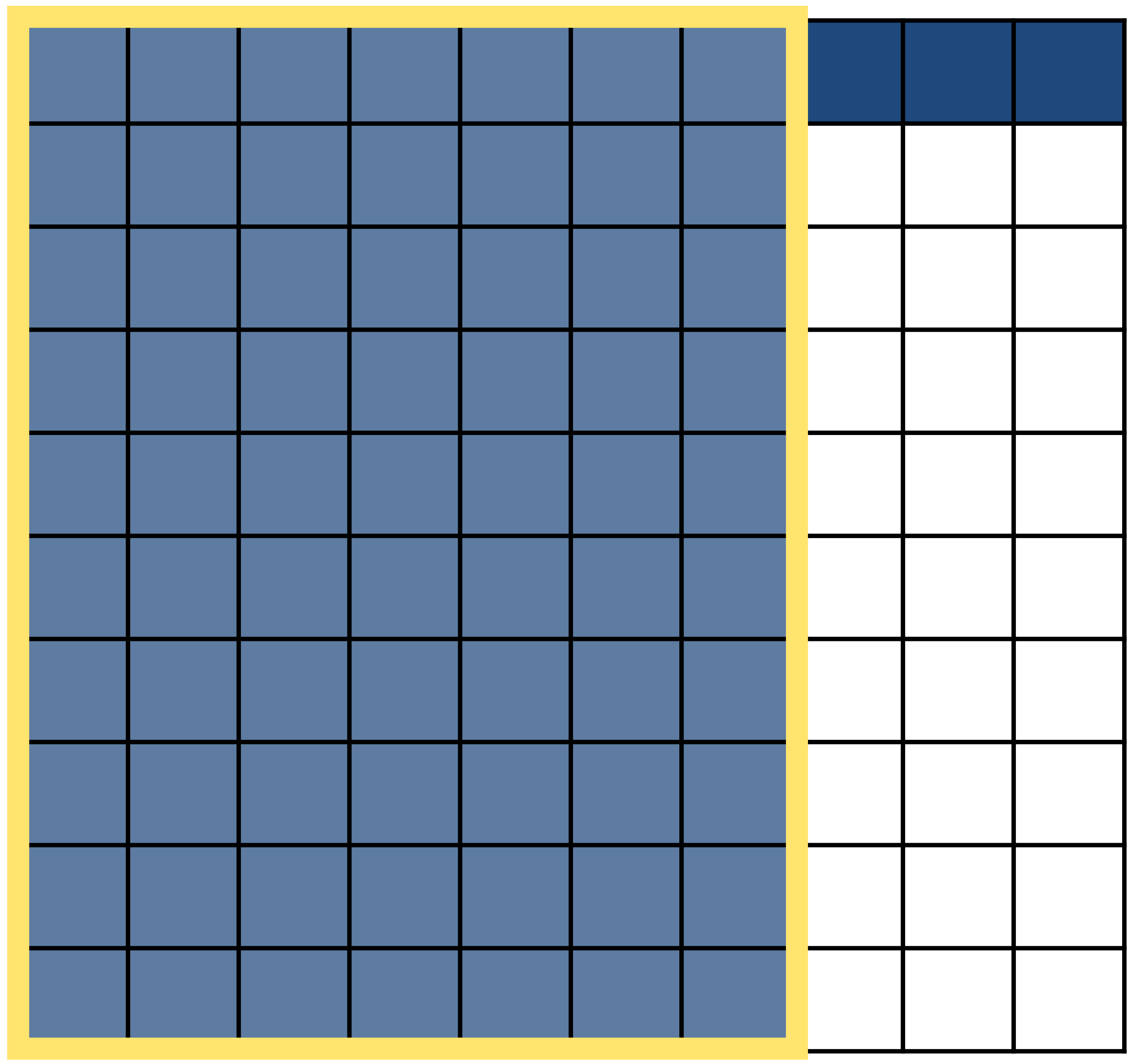
Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures



True distribution

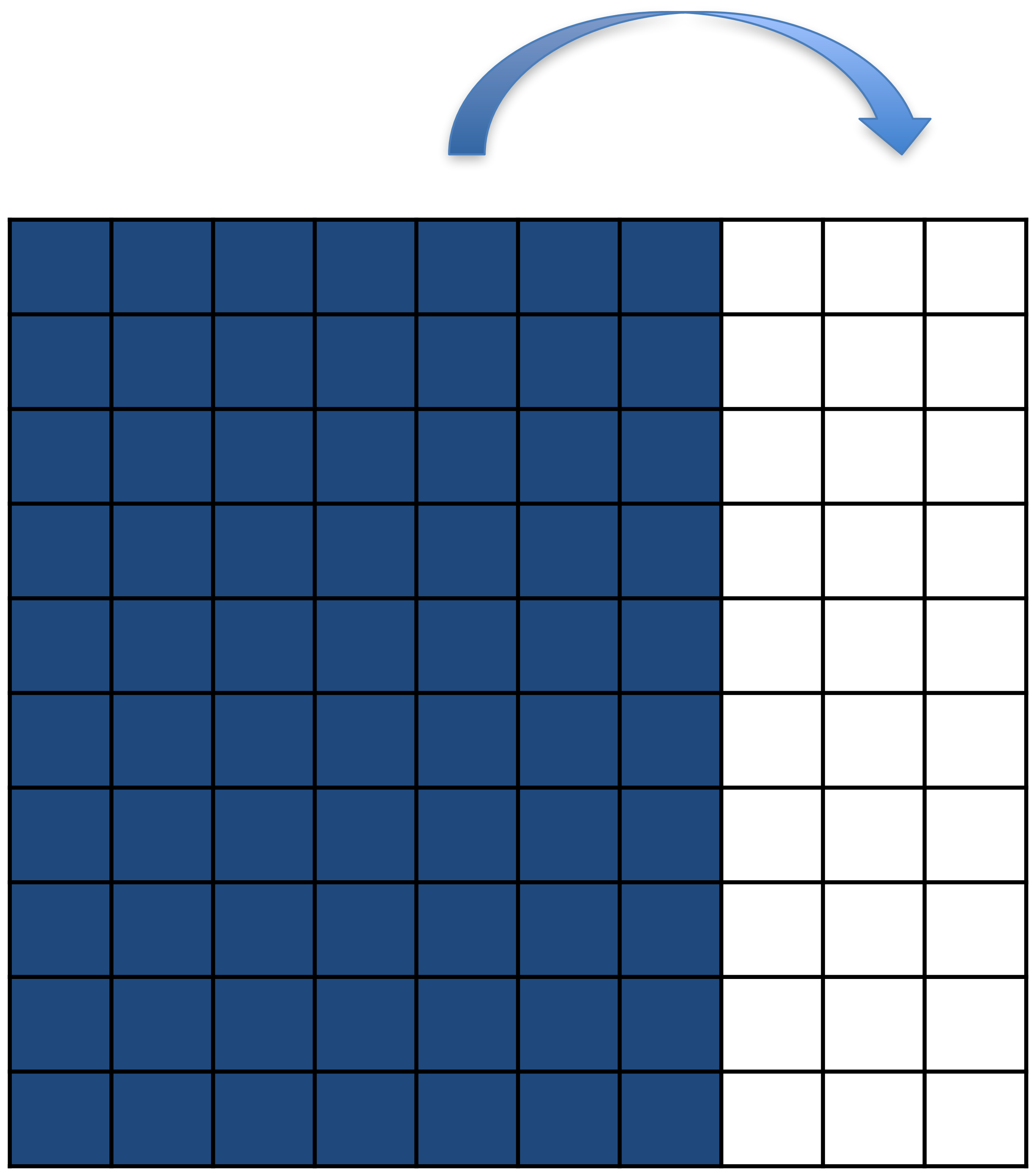
$$\mathrm{P}(\mathrm{supp}(p_D(\mathbf{x}))) \approx 1$$



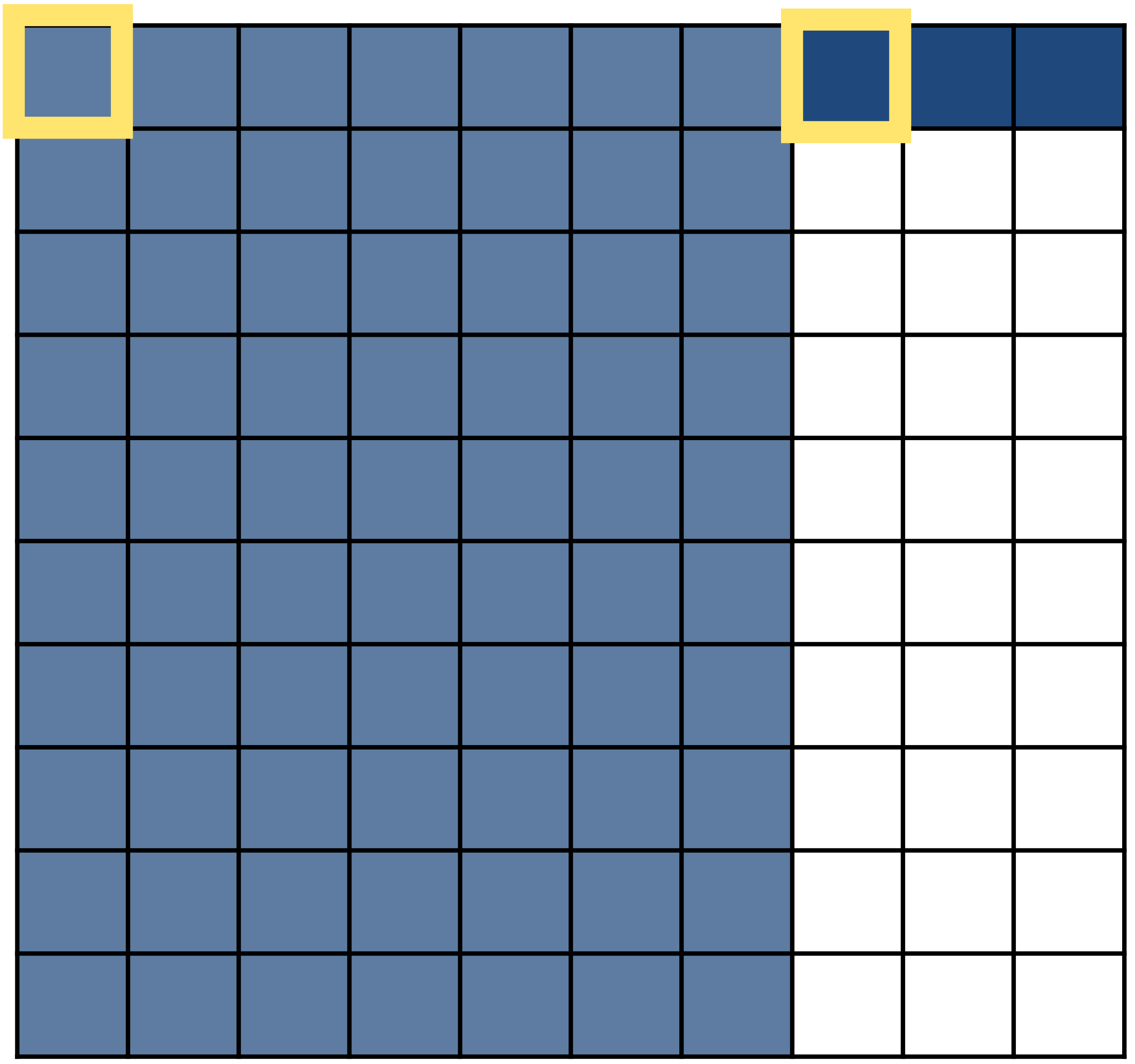
Misestimated model

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures



True distribution



Misestimated model

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures

	True P	$P_{\theta,10^4}$	$P_{\theta,10^3}$	$P_{\theta,10^2}$
LL	-13.8155	-13.8255	-13.8165	-13.8156
$\Pr(\mathrm{supp}(P))$	1.0	0.99	0.999	0.9999
OOD AUC	1.0	0.0	0.0	0.0

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures

	True P	$P_{\theta,10^4}$	$P_{\theta,10^3}$	$P_{\theta,10^2}$
LL	-13.8155	-13.8255	-13.8165	-13.8156
$\Pr(\mathrm{supp}(P))$	1.0	0.99	0.999	0.9999
OOD AUC	1.0	0.0	0.0	0.0

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures

	True P	$P_{\theta,10^4}$	$P_{\theta,10^3}$	$P_{\theta,10^2}$
LL	-13.8155	-13.8255	-13.8165	-13.8156
$\Pr(\mathrm{supp}(P))$	1.0	0.99	0.999	0.9999
OOD AUC	1.0	0.0	0.0	0.0

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

Minimal Estimation Error Can Cause Detection Failures

	True P	$P_{\theta,10^4}$	$P_{\theta,10^3}$	$P_{\theta,10^2}$
LL	-13.8155	-13.8255	-13.8165	-13.8156
$\Pr(\mathrm{supp}(P))$	1.0	0.99	0.999	0.9999
OOD AUC	1.0	0.0	0.0	0.0

Lily H. Zhang, Mark Goldstein, Rajesh Ranganath. "Understanding Failures in Out-of-distribution Detection with Deep Generative Models." *ICML 2021.*

How can we mitigate issues arising from estimation error?

How can we mitigate issues arising from estimation error?

- MLE does not prioritize the bits of information most important for detection:

How can we mitigate issues arising from estimation error?

- MLE does not prioritize the bits of information most important for detection:
  - $\mathbb{E}_{p(x,y)}[-\log p_\theta(x, y)] = \mathbb{E}_{p(y)}\mathbb{E}_{p(x|y)}[-\log p_\theta(x \mid y)] + \mathbb{E}_{p(y)}[-\log p_\theta(y)]$

How can we mitigate issues arising from estimation error?

- MLE does not prioritize the bits of information most important for detection:
  - $\mathbb{E}_{p(x,y)}[-\log p_\theta(x,y)] = \mathbb{E}_{p(y)}\mathbb{E}_{p(x|y)}[-\log p_\theta(x \mid y)] + \mathbb{E}_{p(y)}[-\log p_\theta(y)]$
  - For uniform class distribution, second term is $\log K$, where $K = \#$ classes

How can we mitigate issues arising from estimation error?

- MLE does not prioritize the bits of information most important for detection:
  - $\mathbb{E}_{p(x,y)}[-\log p_\theta(x, y)] = \mathbb{E}_{p(y)}\mathbb{E}_{p(x|y)}[-\log p_\theta(x \mid y)] + \mathbb{E}_{p(y)}[-\log p_\theta(y)]$
  - For uniform class distribution, second term is $\log K$, where $K = \#$ classes
  - Many more bits associated with generating the object

How can we mitigate issues arising from estimation error?

- MLE does not prioritize the bits of information most important for detection:
  - $\mathbb{E}_{p(x,y)}[-\log p_\theta(x, y)] = \mathbb{E}_{p(y)}\mathbb{E}_{p(x|y)}[-\log p_\theta(x \mid y)] + \mathbb{E}_{p(y)}[-\log p_\theta(y)]$
  - For uniform class distribution, second term is $\log K$, where $K =$ # classes
  - Many more bits associated with generating the object
- To prioritize important information in modeling, employ representation learning!

What constitutes a good representation?

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We want representations that can distinguish between classes **y**

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We want representations that can distinguish between classes $\mathbf{y}$
  - $\arg\max_{r} p(\mathbf{y} \mid r(\mathbf{x}))$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want to rely on known spurious signal $\mathbf{z}$ that happens to be correlated with $\mathbf{y}$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want to rely on known spurious signal $\mathbf{z}$ that happens to be correlated with $\mathbf{y}$
  - $\arg\max_r p_{\perp\perp}(\mathbf{y} \mid r(\mathbf{x}))$, where $p_{\perp\perp}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{y}, \mathbf{z})p(\mathbf{y})p(\mathbf{z})$ vs. $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{y}, \mathbf{z})p(\mathbf{y} \mid \mathbf{z})p(\mathbf{z})$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want to rely on known spurious signal $\mathbf{z}$ that happens to be correlated with $\mathbf{y}$
  - $\arg\max_r p_{\perp\!\!\!\perp}(\mathbf{y} \mid r(\mathbf{x}))$, where $p_{\perp\!\!\!\perp}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{y}, \mathbf{z})p(\mathbf{y})p(\mathbf{z})$ vs. $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{y}, \mathbf{z})p(\mathbf{y} \mid \mathbf{z})p(\mathbf{z})$

Approximate via reweighting:

$$p_{\perp\!\!\!\perp}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x}, \mathbf{y}, \mathbf{z})\frac{p(\mathbf{y})}{p(\mathbf{y} \mid \mathbf{z})}$$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

Lily H. Zhang and Rajesh Ranganath. “Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection.” *AAAI 2023*.

What constitutes a good representation?

- We do not want representations that depend on **z** within each class or overall

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want representations that depend on **z** within each class or overall
    - $\arg\max_r p_{\perp\!\!\!\perp}(\mathbf{y} \mid r(\mathbf{x}))$ s.t. $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z}$ and $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} \mid \mathbf{y}$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want representations that depend on $\mathbf{z}$ within each class or overall
  - $\arg\max_{r} p_{\perp\!\!\!\perp}(\mathbf{y} \mid r(\mathbf{x}))$ s.t. $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z}$ and $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} \mid \mathbf{y}$
    - $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} : p_{\perp\!\!\!\perp}(r(\mathbf{x}), \mathbf{z}) = p_{\perp\!\!\!\perp}(r(\mathbf{x}))p_{\perp\!\!\!\perp}(\mathbf{z})$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want representations that depend on $\mathbf{z}$ within each class or overall
  - $\arg\max_r p_{\perp\perp}(\mathbf{y} \mid r(\mathbf{x}))$ s.t. $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\perp}} \mathbf{z}$ and $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\perp}} \mathbf{z} \mid \mathbf{y}$
    - $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\perp}} \mathbf{z} : p_{\perp\perp}(r(\mathbf{x}), \mathbf{z}) = p_{\perp\perp}(r(\mathbf{x}))p_{\perp\perp}(\mathbf{z})$
    - $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\perp}} \mathbf{z} \mid \mathbf{y} : p_{\perp\perp}(r(\mathbf{x}), \mathbf{z} \mid \mathbf{y}) = p_{\perp\perp}(r(\mathbf{x}) \mid \mathbf{y})p_{\perp\perp}(\mathbf{z} \mid \mathbf{y})$

Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

What constitutes a good representation?

- We do not want representations that depend on $\mathbf{z}$ within each class or overall
  - $\arg\max_r p_{\perp\!\!\!\perp}(\mathbf{y} \mid r(\mathbf{x}))$ s.t. $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z}$ and $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} \mid \mathbf{y}$
    - $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} : p_{\perp\!\!\!\perp}(r(\mathbf{x}), \mathbf{z}) = p_{\perp\!\!\!\perp}(r(\mathbf{x}))p_{\perp\!\!\!\perp}(\mathbf{z})$
    - $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} \mid \mathbf{y} : p_{\perp\!\!\!\perp}(r(\mathbf{x}), \mathbf{z} \mid \mathbf{y}) = p_{\perp\!\!\!\perp}(r(\mathbf{x}) \mid \mathbf{y})p_{\perp\!\!\!\perp}(\mathbf{z} \mid \mathbf{y})$

Add a mutual information penalty:

$$\mathrm{I}(r(\mathbf{x}), \mathbf{y}; \mathbf{z}) = \mathbb{E}_{p_{\perp\!\!\!\perp}(r(\mathbf{x}),\mathbf{y},\mathbf{z})} \log \frac{p_{\perp\!\!\!\perp}(r(\mathbf{x}), \mathbf{y}, \mathbf{z})}{p_{\perp\!\!\!\perp}(r(\mathbf{x}), \mathbf{y})p_{\perp\!\!\!\perp}(\mathbf{z})}$$

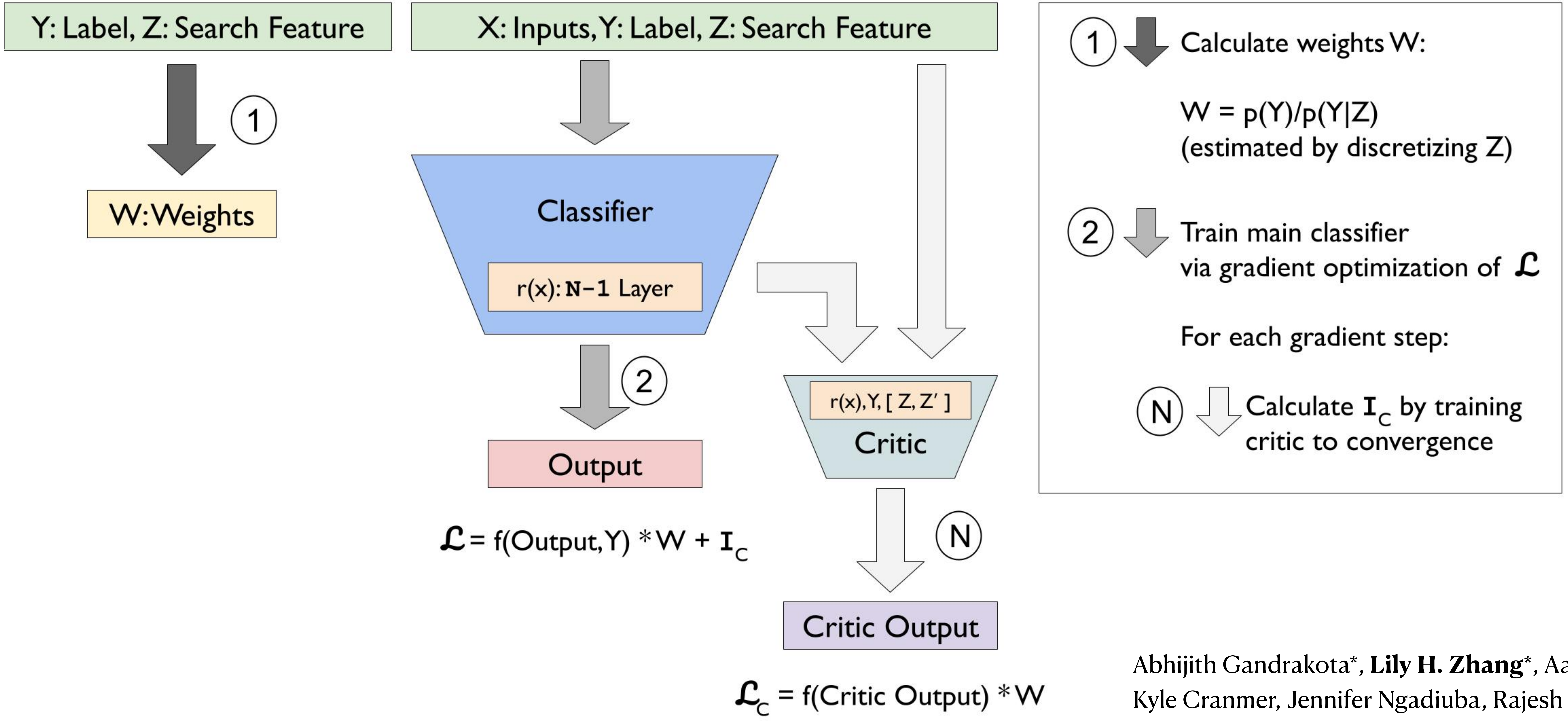
Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

Representation learning for anomaly detection

- We want representations that can distinguish between classes $\mathbf{y}$
  - $\arg\max_r p(\mathbf{y} \mid r(\mathbf{x}))$
- We do not want to rely on known spurious signal $\mathbf{z}$ that happens to be correlated with $\mathbf{y}$
  - $\arg\max_r p_{\perp\!\!\!\perp}(\mathbf{y} \mid r(\mathbf{x}))$
- We do not want representations correlated with $\mathbf{z}$ within each class or overall
  - $\arg\max_r p_{\perp\!\!\!\perp}(\mathbf{y} \mid r(\mathbf{x}))$ s.t. $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z}$ and $r(\mathbf{x}) \perp\!\!\!\perp_{p_{\perp\!\!\!\perp}} \mathbf{z} \mid \mathbf{y}$

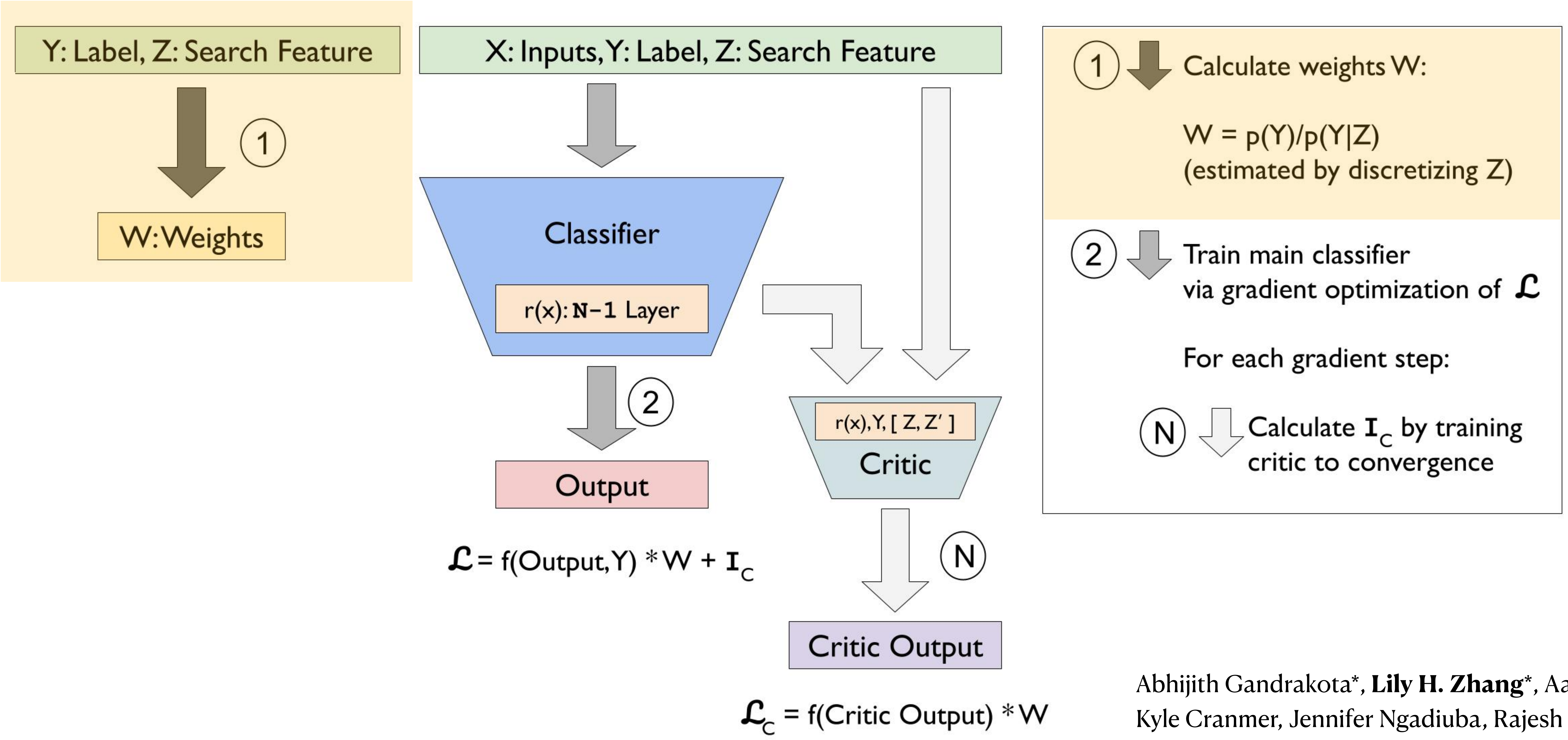
Lily H. Zhang and Rajesh Ranganath. "Robustness to Spurious Correlations Improves Semantic Out-of-distribution Detection." *AAAI 2023*.

Representation learning for anomaly detection



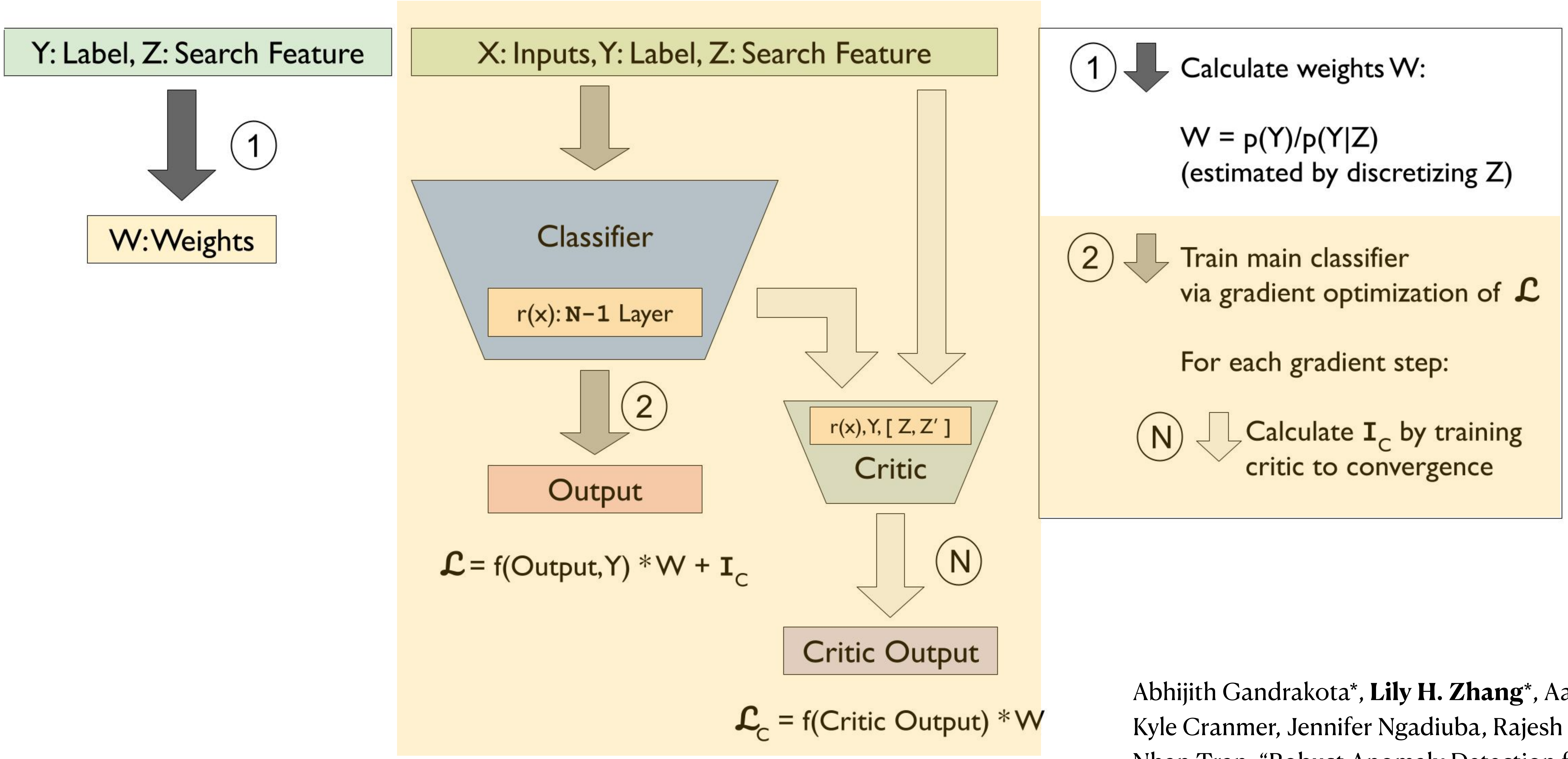
Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Representation learning for anomaly detection



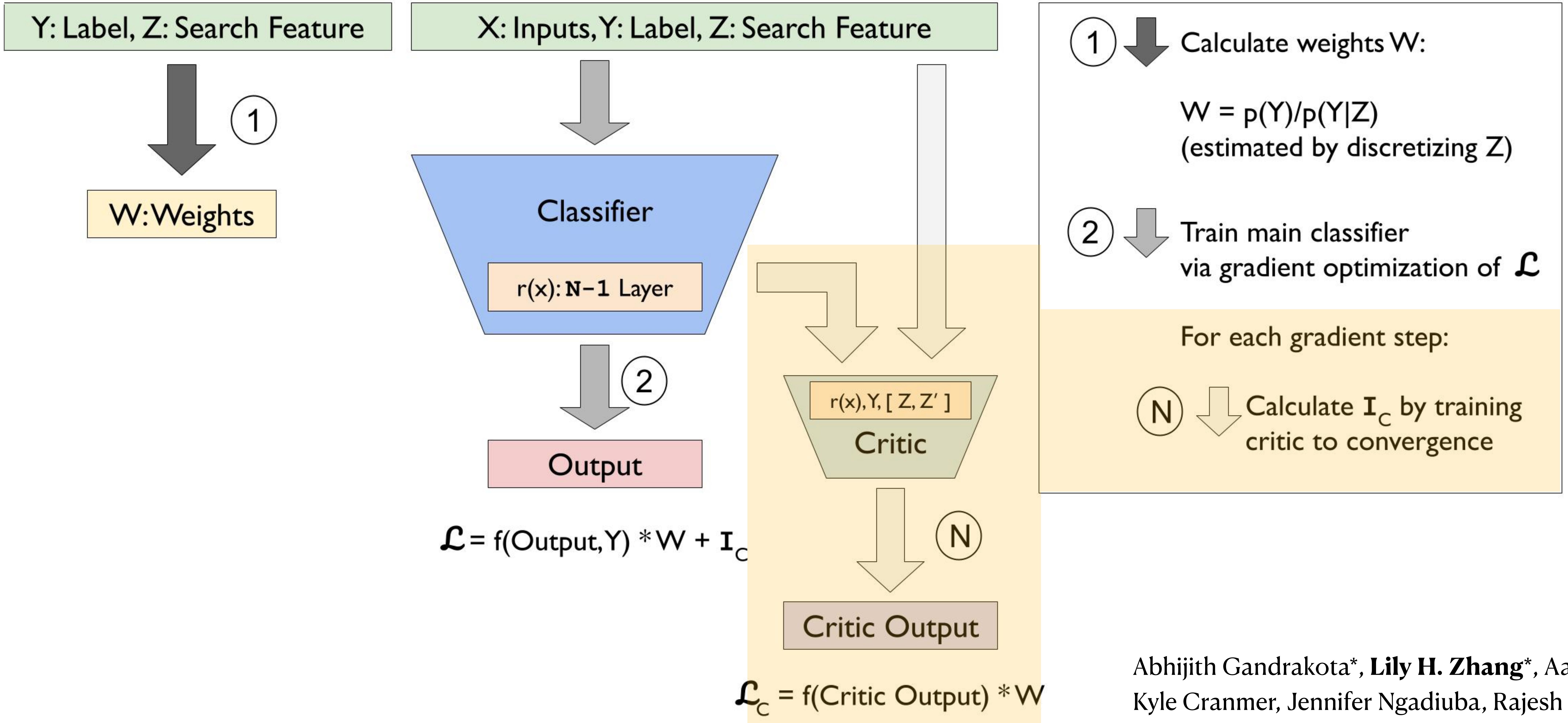
Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Representation learning for anomaly detection



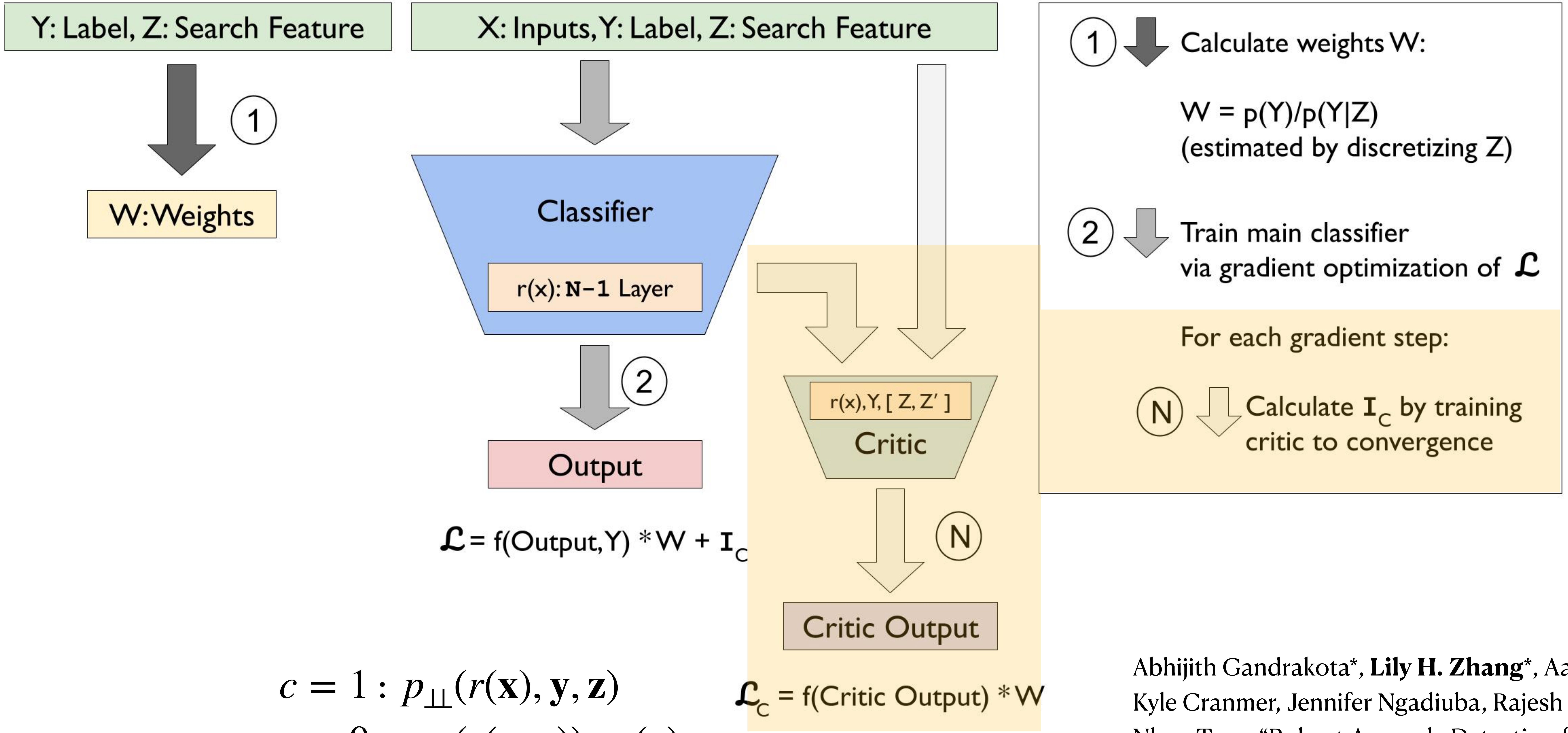
Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Representation learning for anomaly detection



Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

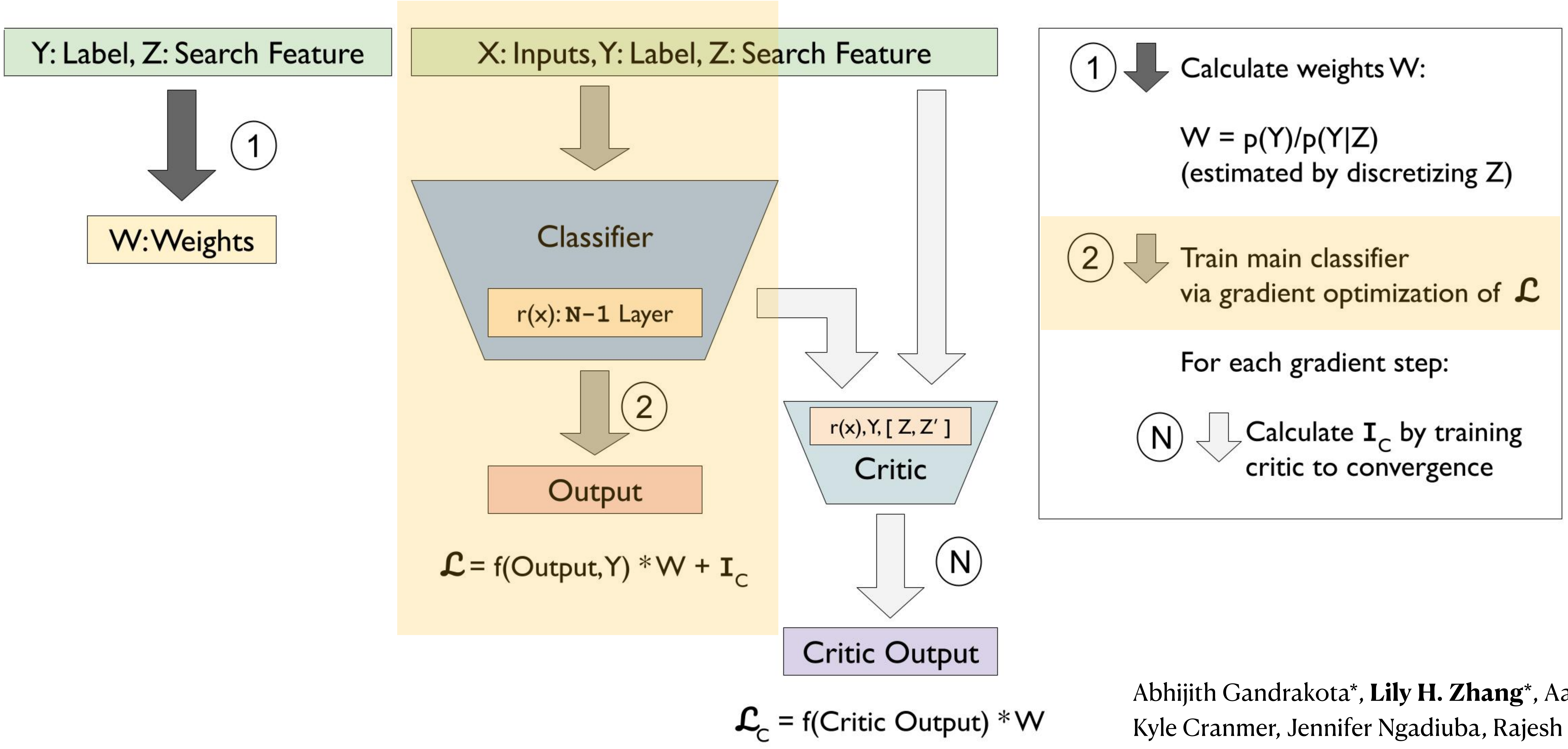
Representation learning for anomaly detection



$$c = 1: \; p_{\perp\!\!\!\perp}(r(\mathbf{x}), \mathbf{y}, \mathbf{z})$$
$$c = 0: \; p_{\perp\!\!\!\perp}(r(\mathbf{x}, \mathbf{y}))p_{\perp\!\!\!\perp}(\mathbf{z})$$

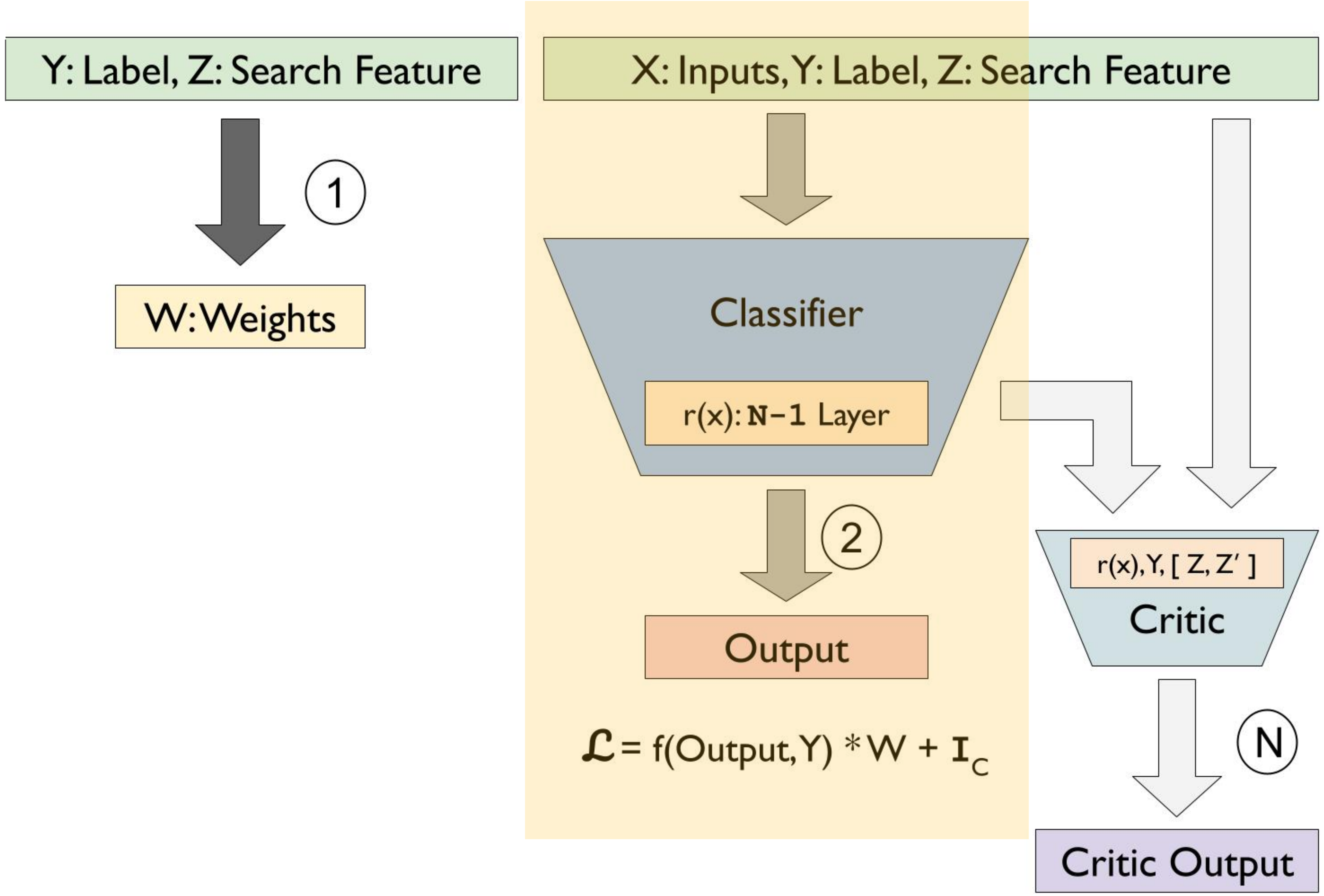
Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Representation learning for anomaly detection



Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Representation learning for anomaly detection



1 Calculate weights W:

W = p(Y)/p(Y|Z)
(estimated by discretizing Z)

2 Train main classifier via gradient optimization of $\mathcal{L}$

For each gradient step:

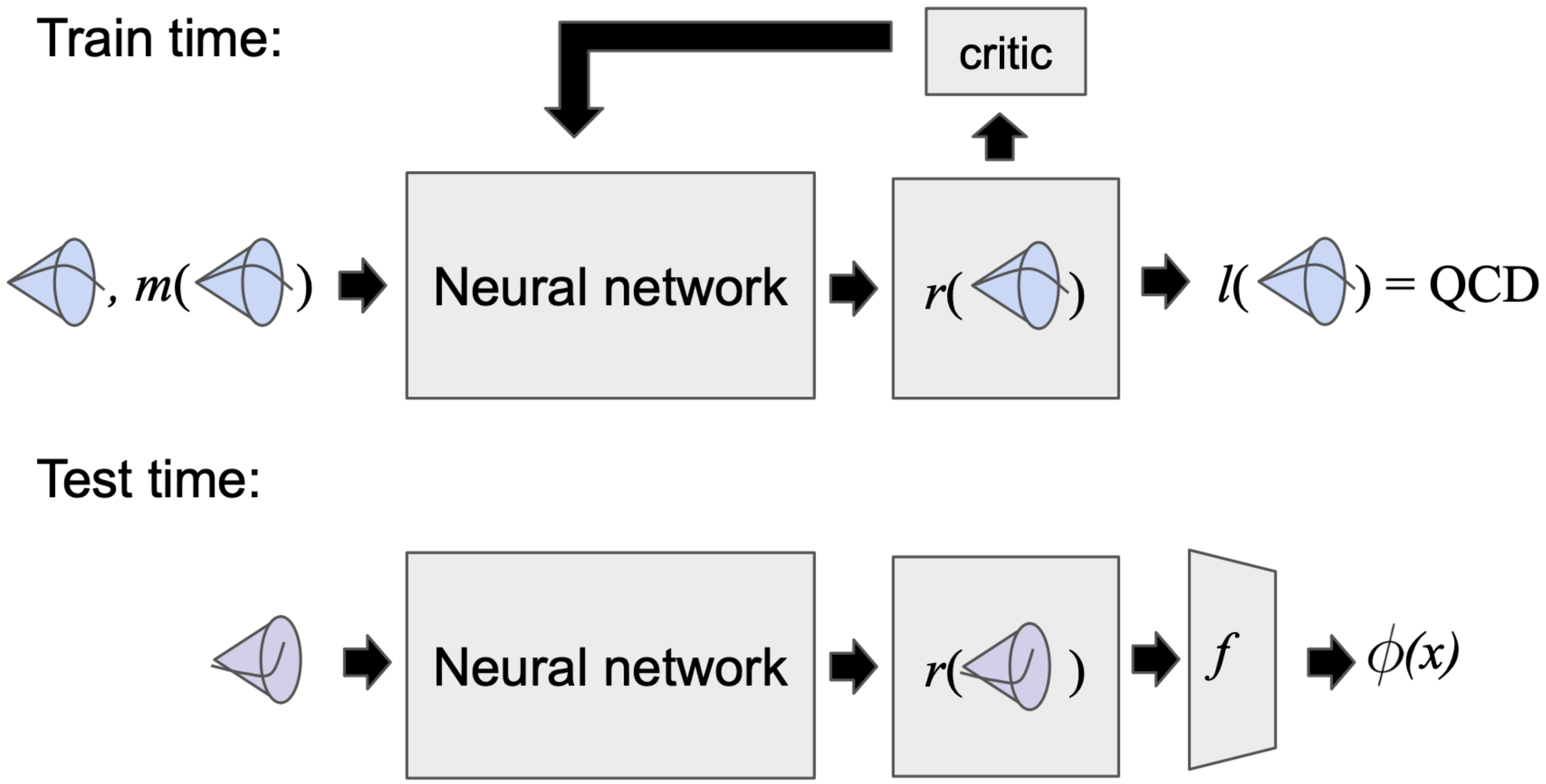
N Calculate $\mathbf{I}_C$ by training critic to convergence

$$I_c = \mathbb{E}_{p_{\perp\!\!\!\perp}(r(\mathbf{x}),\mathbf{y},\mathbf{z})}[\log p_\gamma(c=1 \mid r(\mathbf{x}),\mathbf{y},\mathbf{z}) - \log p_\gamma(c=0 \mid r(\mathbf{x}),\mathbf{y},\mathbf{z})]$$

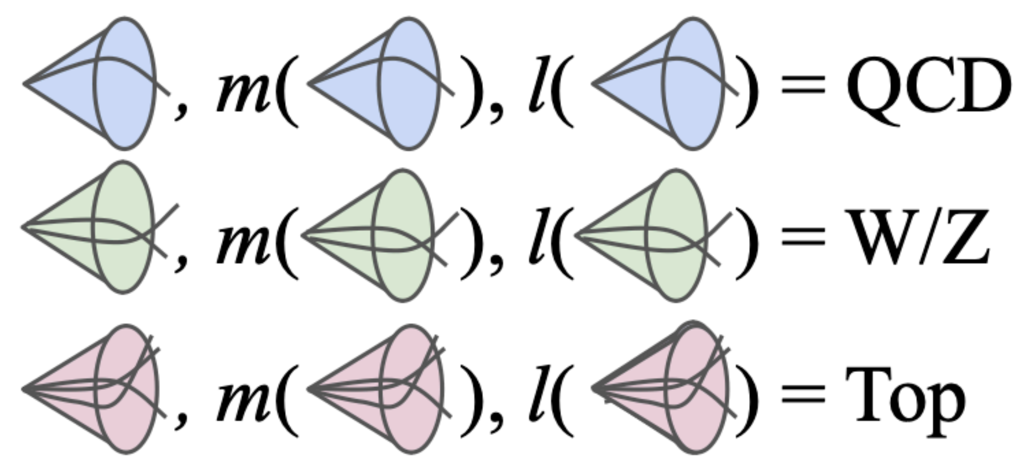
$\mathcal{L}_C$ = f(Critic Output) * W

Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Multi-Background and Nuisance-Aware Representation Learning



Data:



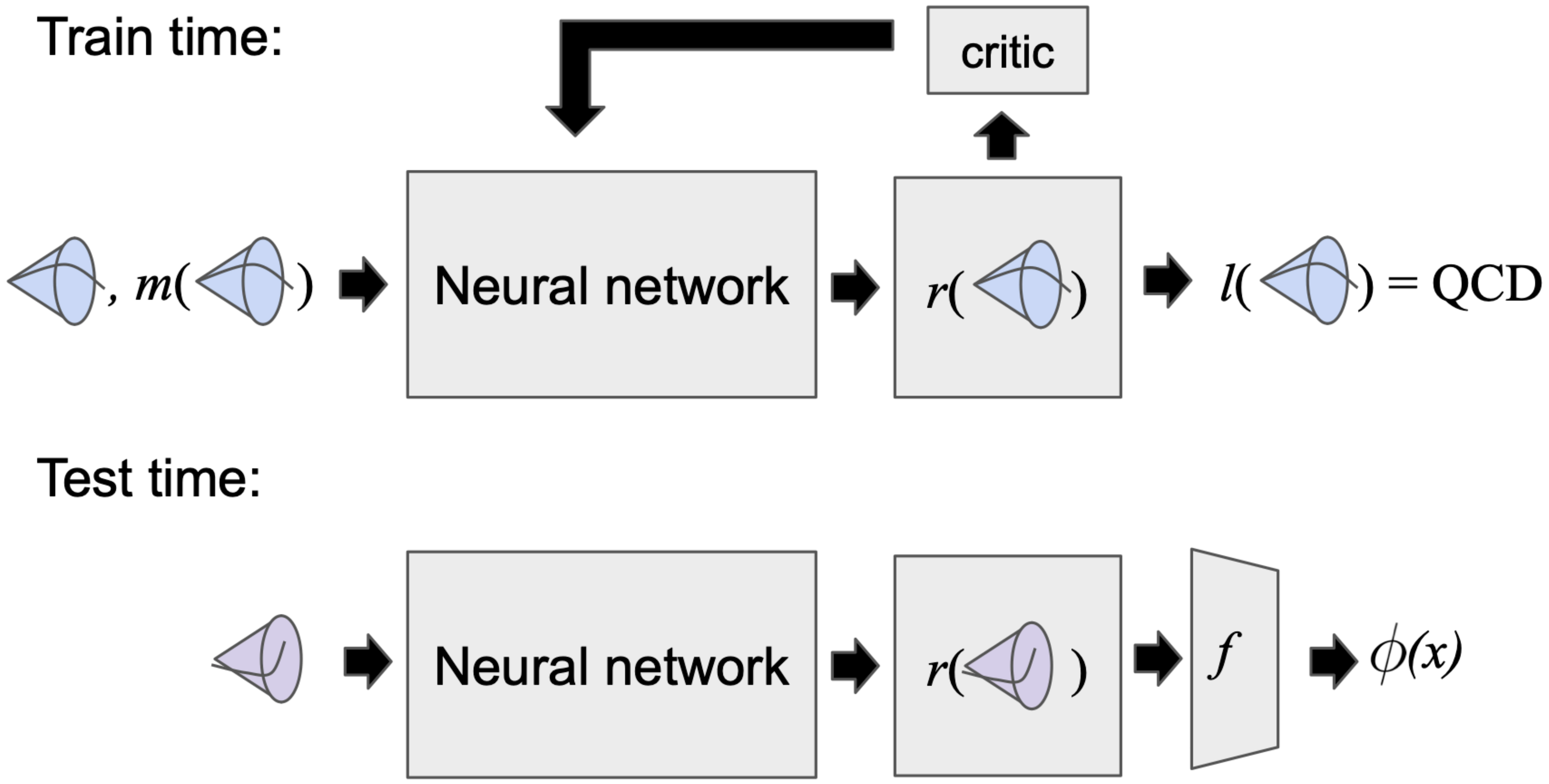
Anomaly score:

f_{ML} = maximum logit

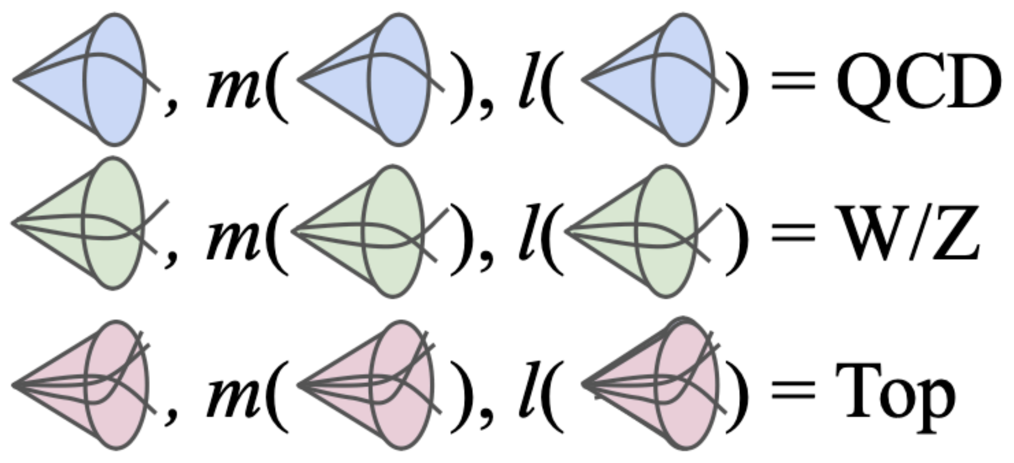
f_{MD} = Mahalanobis distance

Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Multi-Background and Nuisance-Aware Representation Learning



Data:



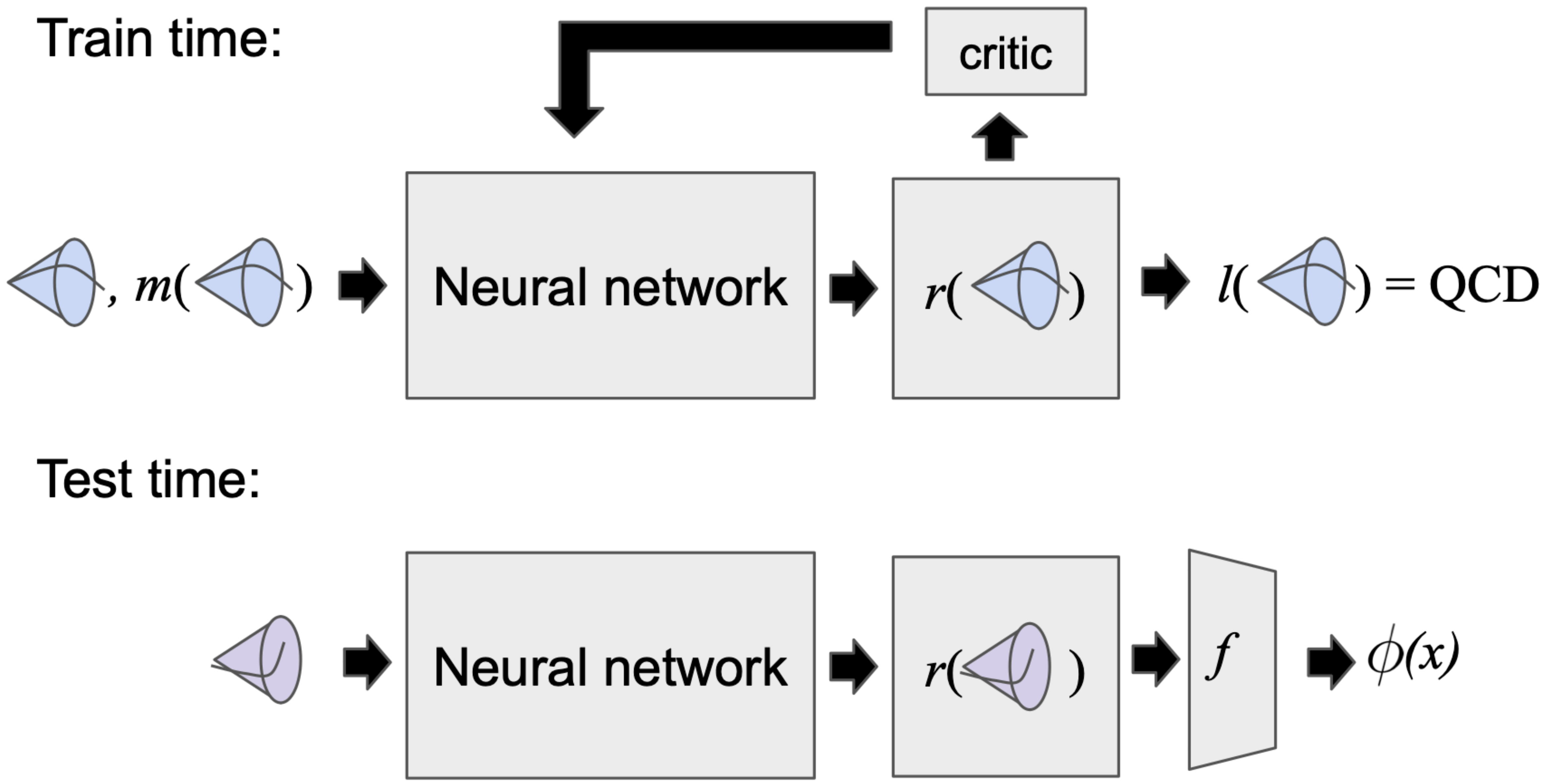
Anomaly score:

f_{ML} = maximum logit

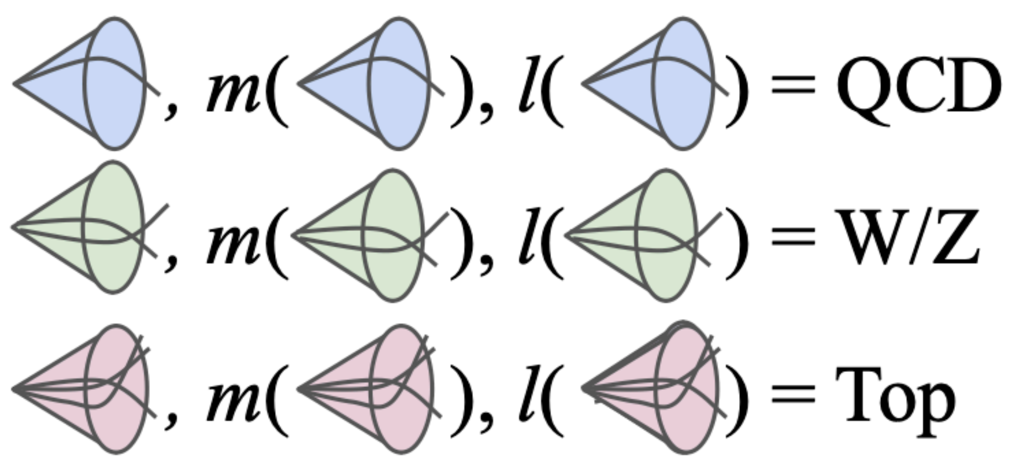
f_{MD} = Mahalanobis distance

Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. "Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning." *MLST 2024*.

Multi-Background and Nuisance-Aware Representation Learning



Data:



Anomaly score:

f_{ML} = maximum logit

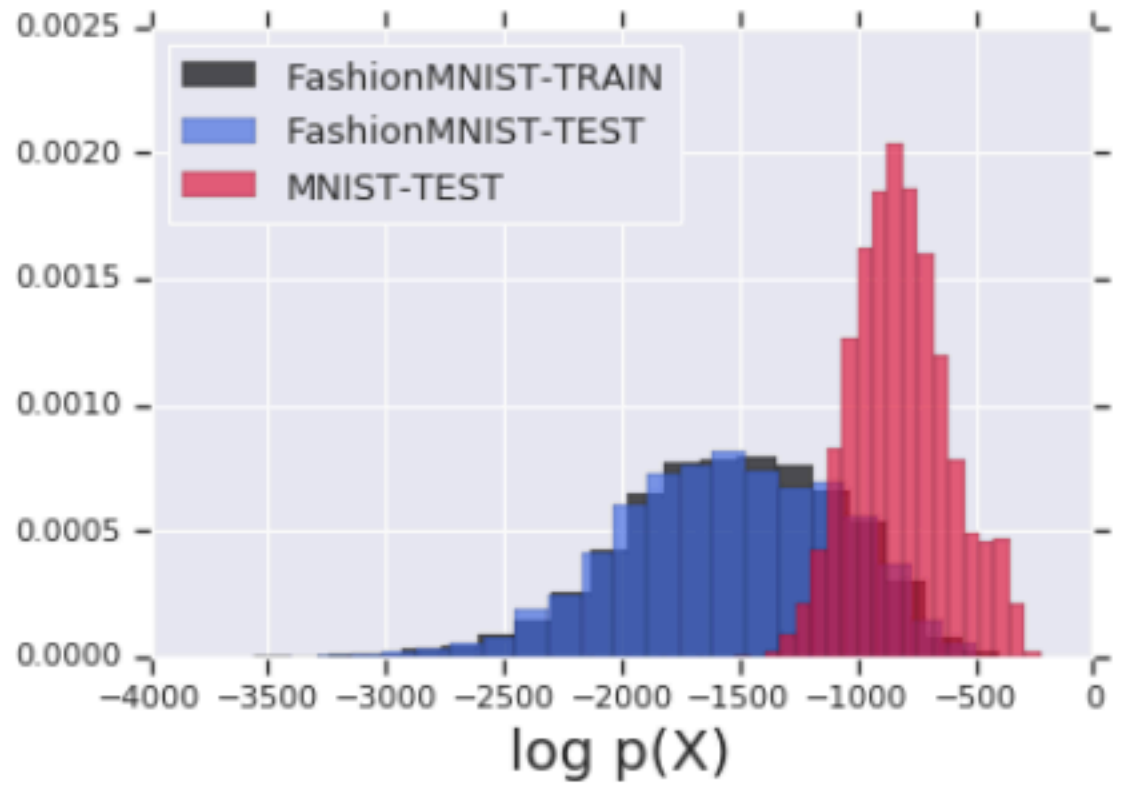
f_{MD} = Mahalanobis distance

Method	AUROC (↑)	JSD (↓)	L2 WD (↓)	SI (↑)
VAE	0.881	0.255	34.3	2.03
nurd-ml	**0.914**	**0.168**	**24.4**	**2.32**
nurd-md	**0.884**	**0.118**	**19.1**	**2.23**

Abhijith Gandrakota*, **Lily H. Zhang***, Aahlad Puli, Kyle Cranmer, Jennifer Ngadiuba, Rajesh Ranganath, Nhan Tran. “Robust Anomaly Detection for Particle Physics using Multi-Background Representation Learning.” *MLST 2024*.

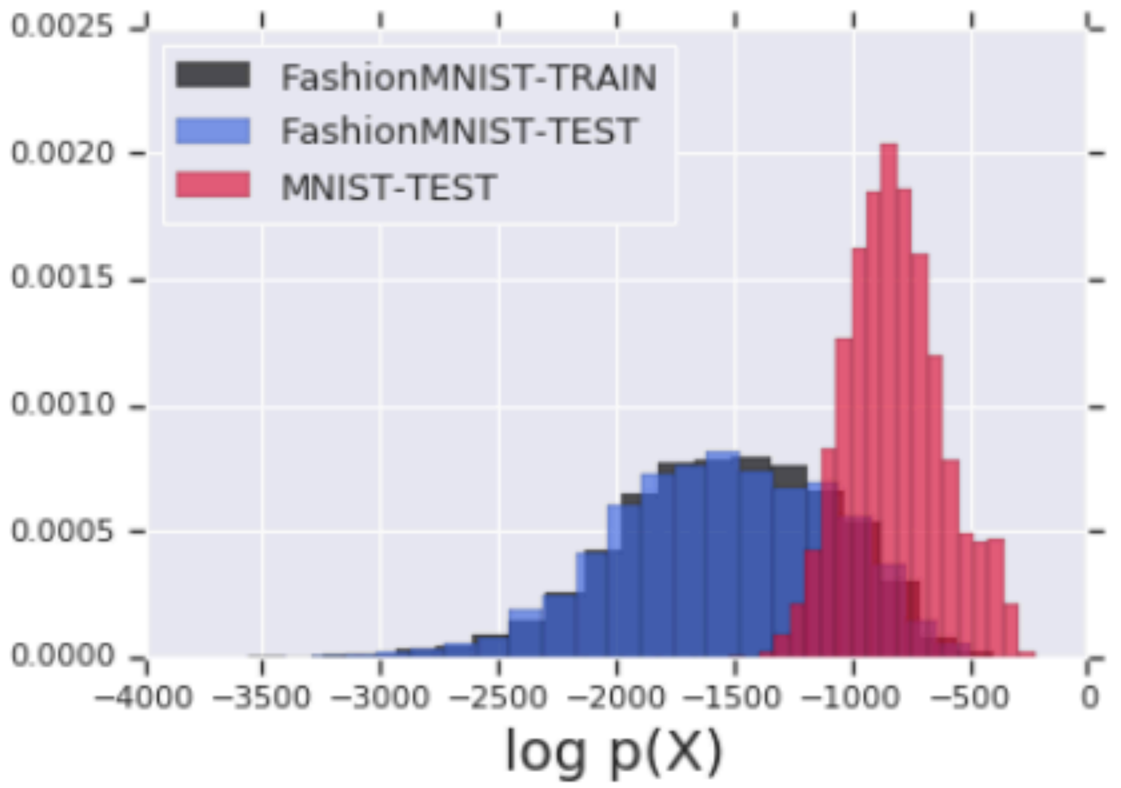
Takeaways

Takeaways

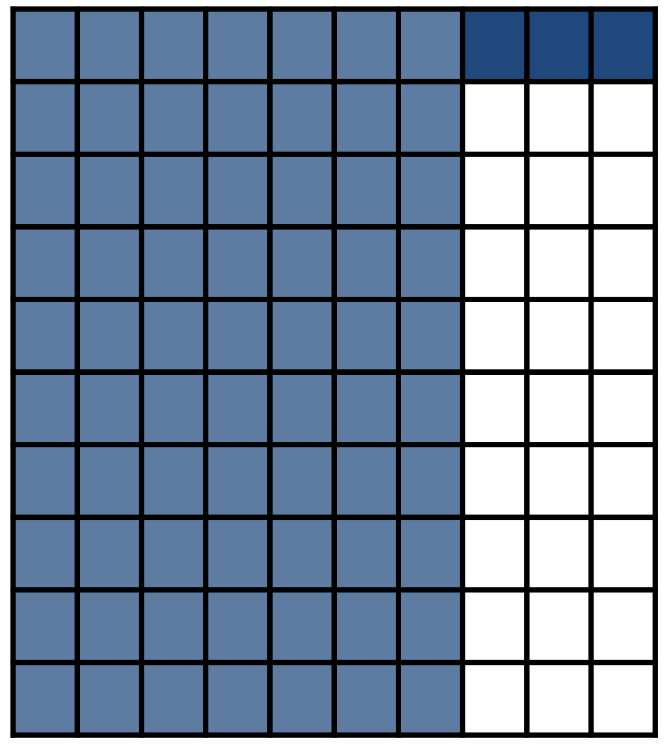


Generative models exhibit detection failures.

Takeaways

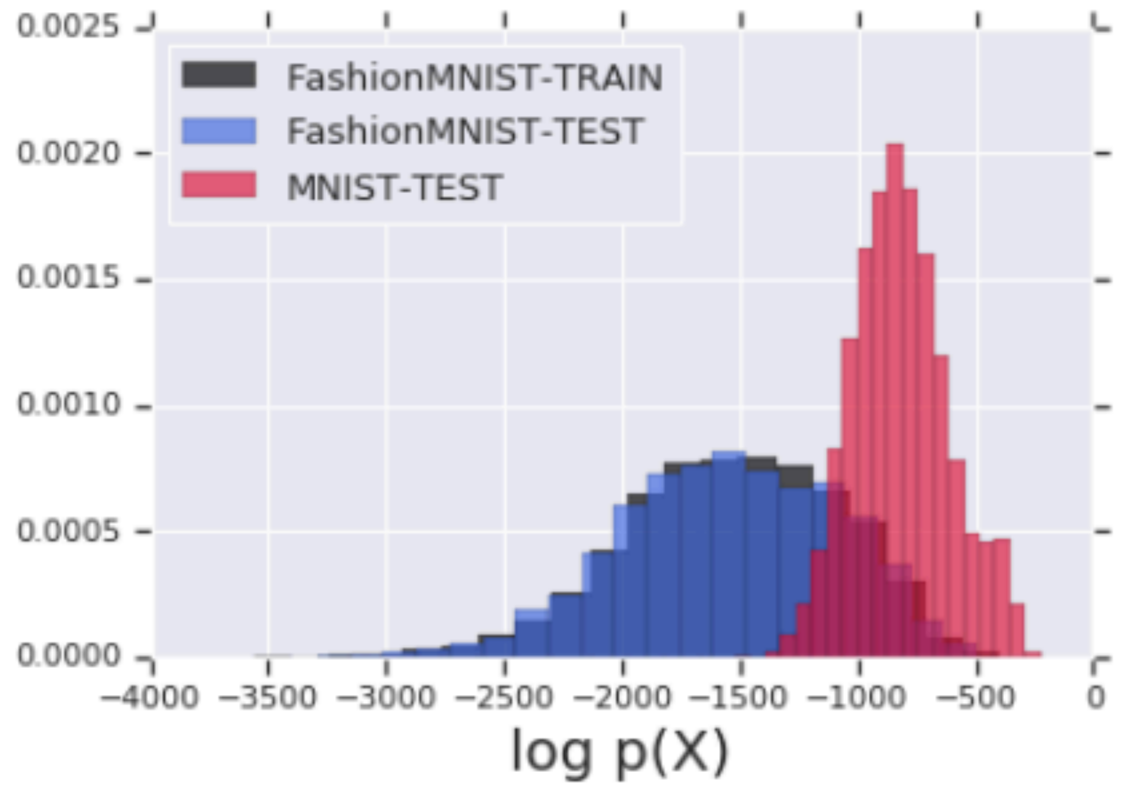


Generative models exhibit detection failures.

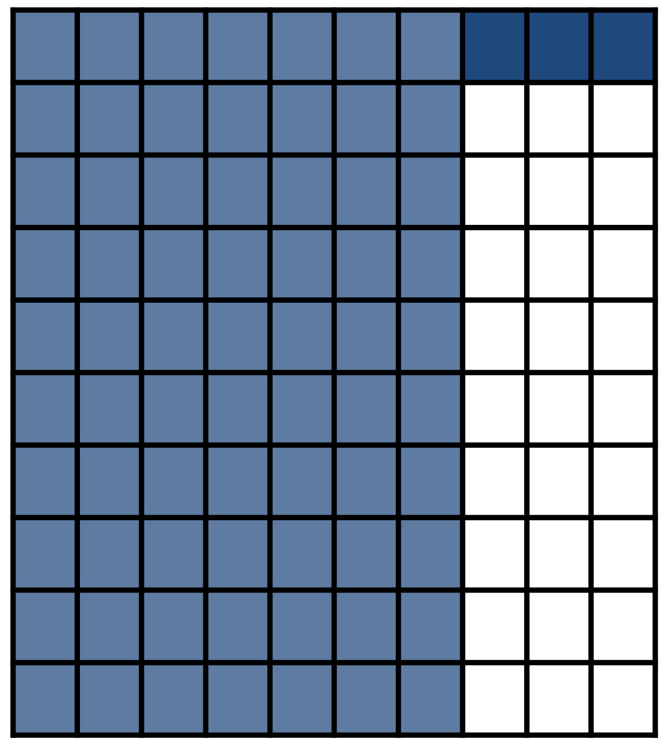


Failures can result from even minimal estimation error.

Takeaways



Generative models exhibit detection failures.

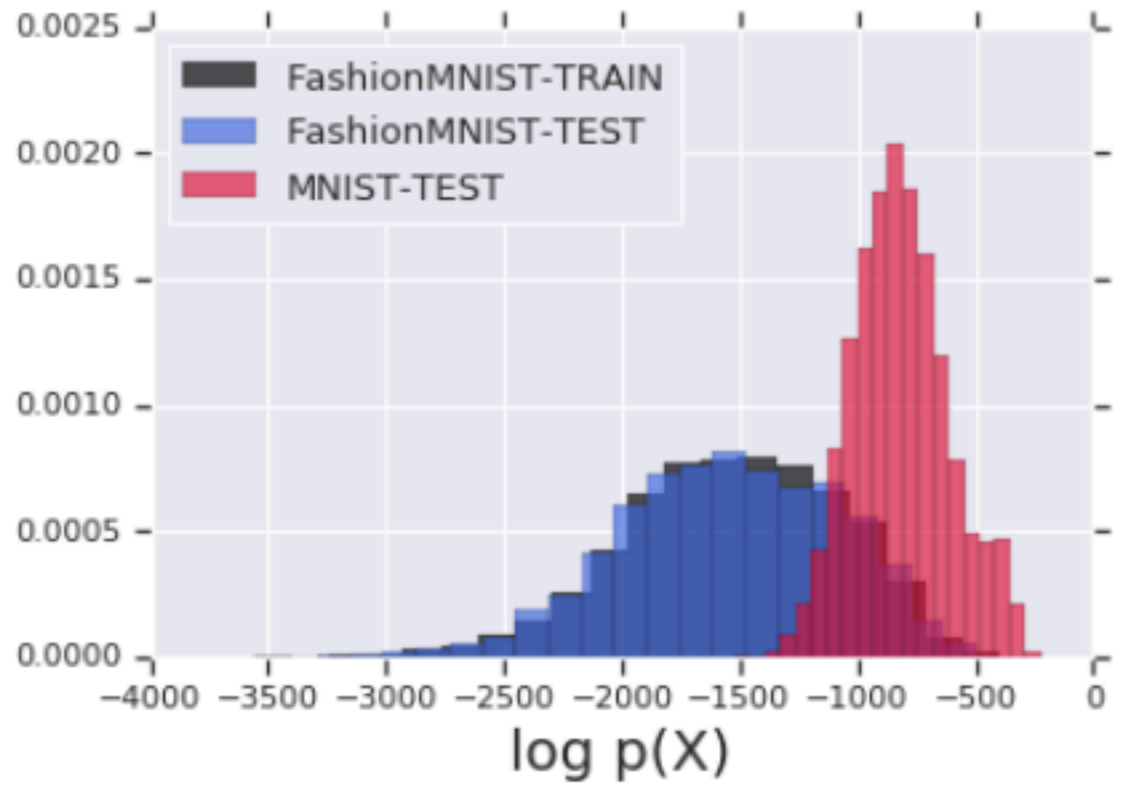


Failures can result from even minimal estimation error.

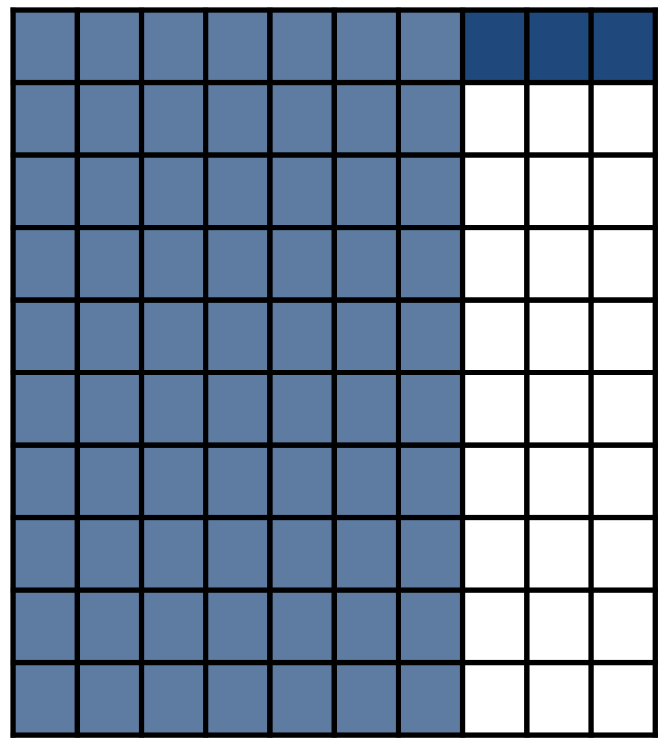
$$\phi_{\text{typical}}(\mathbf{x}) = -\left|\log p_\theta(\mathbf{x}) - \frac{1}{n}\sum_{\mathbf{x}'\in D_{tr}} \log p_\theta(\mathbf{x}')\right|$$

Alternative test statistics can correct for estimation error.

Takeaways



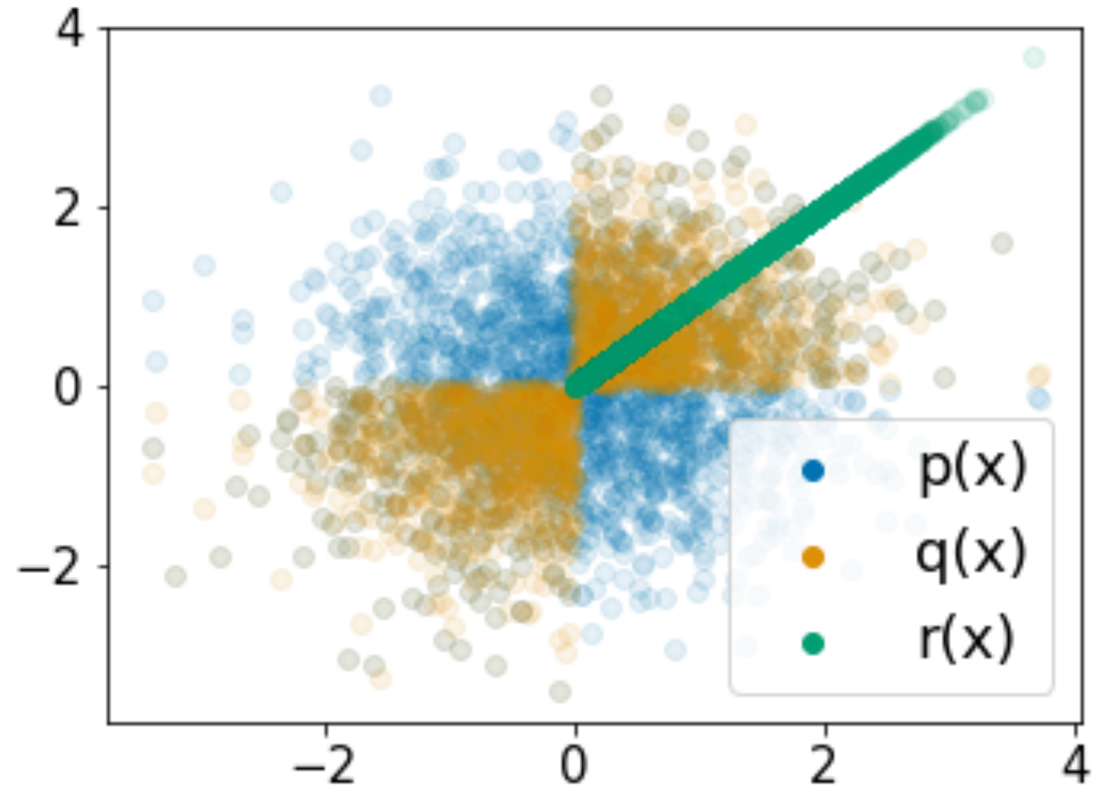
Generative models exhibit detection failures.



Failures can result from even minimal estimation error.

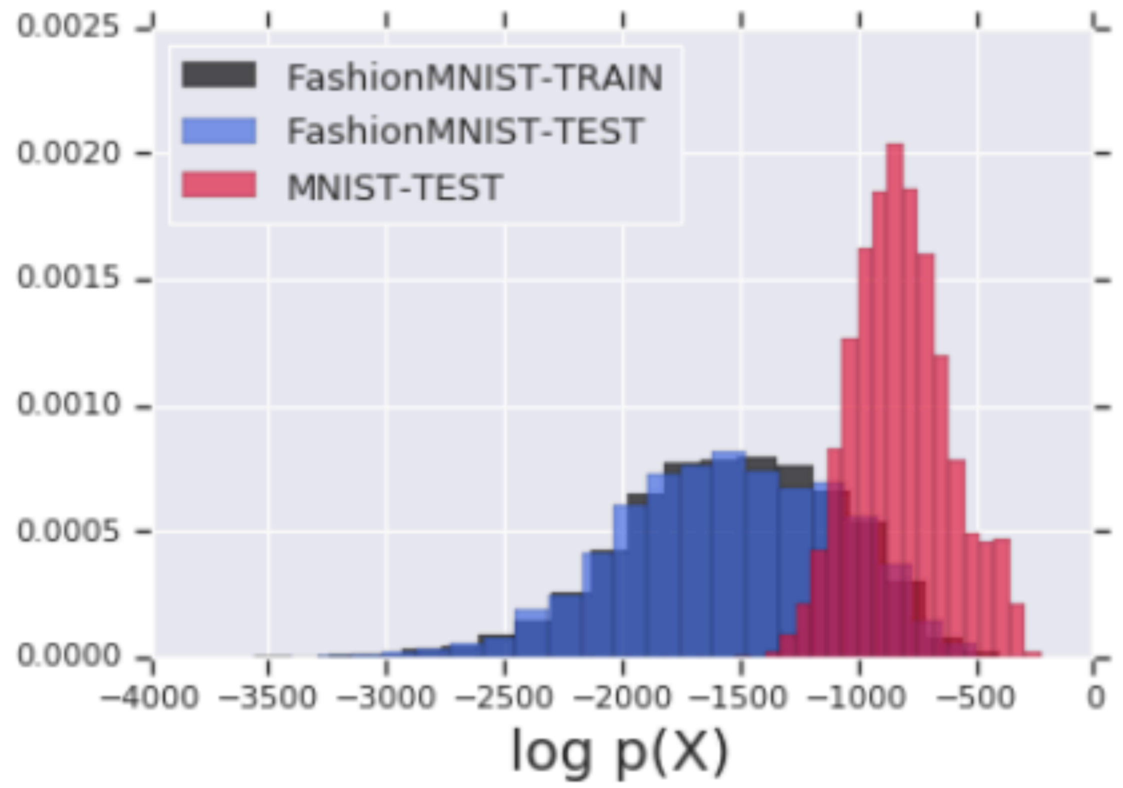
$$\phi_{\text{typical}}(\mathbf{x}) = -\left|\log p_\theta(\mathbf{x}) - \frac{1}{n}\sum_{\mathbf{x}'\in D_{tr}} \log p_\theta(\mathbf{x}')\right|$$

Alternative test statistics can correct for estimation error.

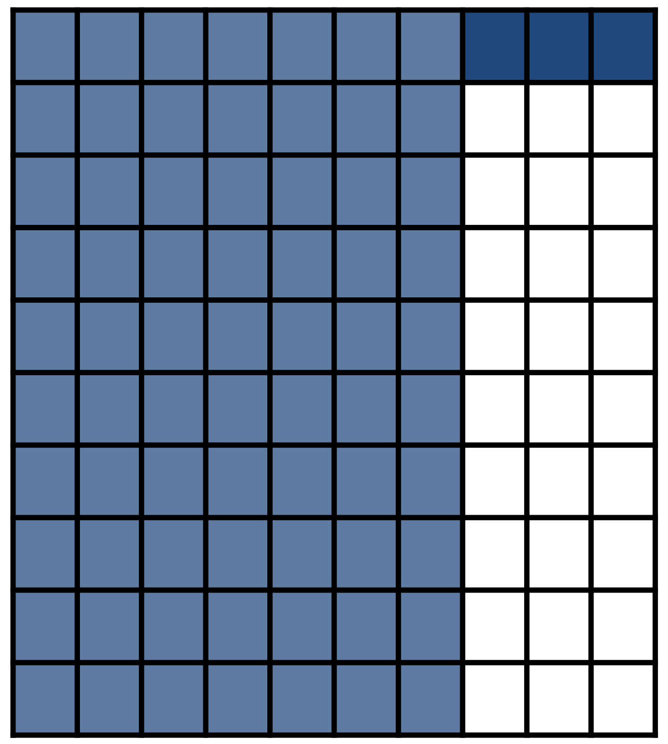


The "right" method depends on assumptions on out-distributions.

Takeaways



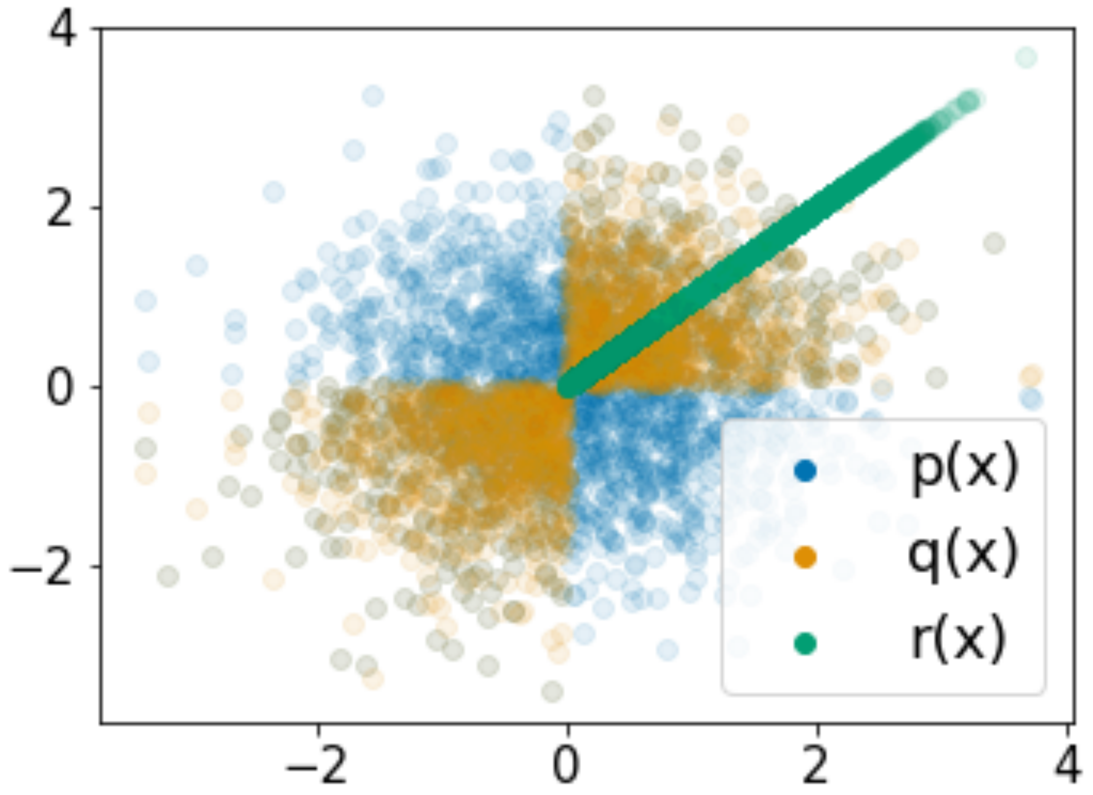
Generative models exhibit detection failures.



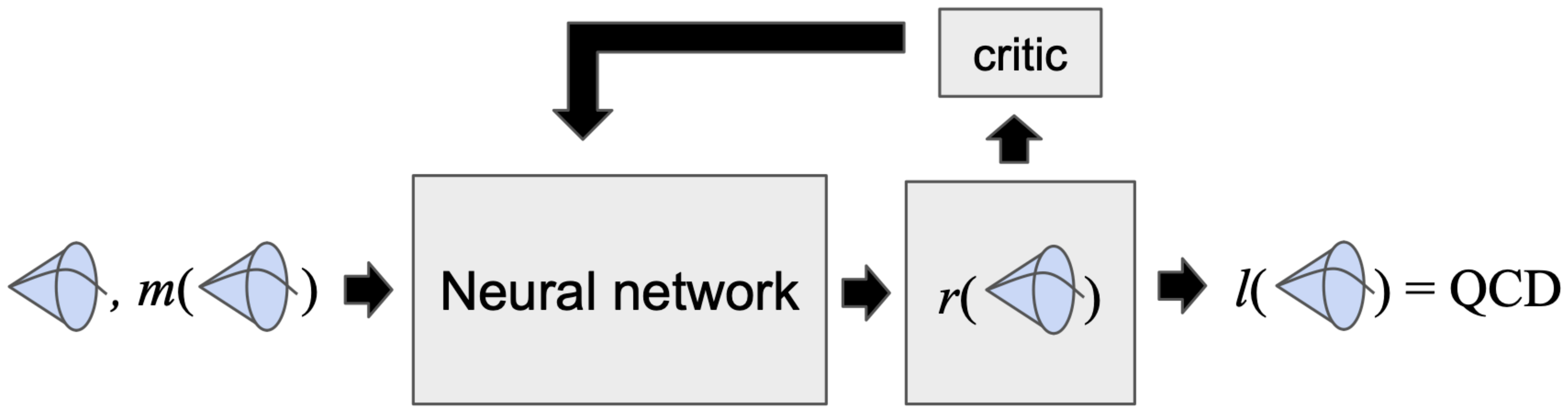
Failures can result from even minimal estimation error.

$$\phi_{\text{typical}}(\mathbf{x}) = -\left| \log p_\theta(\mathbf{x}) - \frac{1}{n} \sum_{\mathbf{x}' \in D_{tr}} \log p_\theta(\mathbf{x}') \right|$$

Alternative test statistics can correct for estimation error.



The “right” method depends on assumptions on out-distributions.



Rather than rely entirely on generative models, consider learning good representations.